Spring 2021

Exposure assessment of emerging contaminants: Rapid screening and modeling of plant uptake

Majid Bagheri

Department: Civil, Architectural and Environmental Engineering

EXPOSURE ASSESSMENT OF EMERGING CONTAMINANTS: RAPID SCREENING AND MODELING OF PLANT UPTAKE

by

MAJID BAGHERI

A DISSERTATION

Presented to the Graduate Faculty of the

MISSOURI UNIVERSITY OF SCIENCE AND TECHNOLOGY

In Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

in

CIVIL ENGINEERING

2021

Approved by:

Joel G. Burken, Advisor
Honglan Shi
Jianmin Wang
Donald C. Wunsch
Lorenzo Rossi

PUBLICATION DISSERTATION OPTION

This dissertation consists of the following three articles, formatted in the style used by the Missouri University of Science and Technology:

Paper I, found on pages 18–46, has been published in SCIENCE OF THE TOTAL ENVIRONMENT *Journal*.

Paper II, found on pages 47–79, has been published in SCIENCE OF THE TOTAL ENVIRONMENT *Journal*.

Paper III, found on pages 80–115, has been published in SCIENCE OF THE TOTAL ENVIRONMENT *Journal*.

ABSTRACT

With the advent of new chemicals and their increasing uses in every aspect of our life, considerable number of emerging contaminants are introduced to environment yearly. Emerging contaminants in forms of pharmaceuticals, detergents, biosolids, and reclaimed wastewater can cross plant roots and translocate to various parts of the plants. Long-term human exposure to emerging contaminants through food consumption is assumed to be a pathway of interest. Thus, uptake and translocation of emerging contaminants in plants are important for the assessment of health risks associated with human exposure to emerging contaminants. To have a better understanding over fate of emerging contaminants in plants, both experimental and modeling studies are necessary. In our studies, various machine learning techniques were used to develop models with improved accuracies as compared to traditional models. Our results showed that neural network models can considerably improve the accuracy of predictions for transpiration stream concentration factor (TSCF) and root concentration factor (RCF). With fuzzy logic as another capable machine learning technique, it was possible to develop new relationships between chemical properties and uptake efficiency. In our experimental studies, the uptake of various organic compounds including estriol, bisphenol A, carbamazepine, DEET, lincomycin, acetaminophen, and 2,4-dinitrotoluene was in a good agreement with predictive models. The moderately hydrophobic compounds had the highest TSCF values. However, our findings indicated that the role of factors other than physicochemical properties should be taken into account. Degradation of chemical compounds and partitioning to plant root constituents other than lipids (lignin, cellulose, and protein) are some examples.

ACKNOWLEDGMENTS

I would like to express special thanks of gratitude to my distinguished advisor Dr. Joel G. Burken who gave me the golden opportunity to join his wonderful research group at Missouri University of Science and Technology. I would also like to thank my respected co-advisor Dr. Honglan Shi who always supported me during my PhD. Both my advisor and co-advisor helped me in doing a lot of novel research, and I came to know about so many new things. Their advice and guidance not only helped me to solve problems but also enhanced my confidence in dealing with research problems at cutting-edge.

I would also like to thank Dr. Donlad C. Wunsch, Dr. Jianmin Wang, and Dr. Lorenzo Rossi who kindly agreed to serve on my committee. Their comments and concerns were very valuable and helped me to improve the quality of conducted studies. In addition, working with them during this long journey gave me the opportunity to learn about different research and teaching styles.

Last but not least, I would like to thank my family and friends who helped and supported me in finalizing this project within the limited time frame. With the love and support of my parents and other family members, I could overcome all the challenges and did my best to achieve the intended goals.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

PAPER I Page

PAPER II

PAPER III

LIST OF TABLES

NOMENCLATURE

Symbol	Description
φ	Activation Function
f	Transfer Function

1. INTRODUCTION

1.1. UPTAKE OF CONTAMINANTS BY PLANTS

With the advent of new chemicals and their growing uses for different purposes, most of them and their derivatives end up in soil and water (Teng et al., 2014). These chemical compounds can interact with living organisms, and plants as primary producers are intentionally or even unintentionally exposed to increasing number of these novel and new chemical compounds called emerging contaminants (Pullagurala et al., 2018). The emerging contaminants can cross plant roots and translocate to various parts of the plants including edible parts. Uptake and translocation of emerging contaminants is complicated and affected by several factors. Thus, the potential impact of plants on the uptake and overall translocation of emerging contaminants is worth detailed scrutiny (Picó et al., 2019; Pullagurala et al., 2018). Physicochemical properties of chemicals have always been considered as driving factors in the overall uptake and translocation of emerging contaminants from subsurface into the plants. Lipophilicity as measured by octanol/water partitioning (log K_{ow}) is one of the most important physicochemical properties. Multiple bodies of work have demonstrated K_{ow} to be the dominant factor the partitioning of chemicals between water and plant roots (Burken and Schnoor, 1998; Briggs et al., 1982; Dettenmaier et al., 2009; Shone et al., 1974). In contrast to lipophilicity, limited understanding has been communicated about the role and importance of other physicochemical properties and the development of poly-parameter models is limiting. In addition to these properties, the impacts of factors related to soil and plants should also be taken into account. The soil properties such as pH can affect the interactions between

chemicals and plant roots. Partitioning of chemicals between water and plant roots is also affected by fundamental materials of plant roots such as lignin, cellulose, and protein. To have improved understanding over the role of these factors in the uptake and translocation of emerging contaminants in plants, it is necessary to conduct uptake and partitioning experiments in which chemicals cover a wide range of physicochemical properties.

1.2. MEAUREMENT OF PLANT UPTAKE

To measure the uptake and translocation of environmental contaminants in plants, various concentration factors have been developed. These factors compare the concentration of chemicals in plant tissues to the concentration of that chemicals in the subsurface. Root concentration factor (RCF) is a ratio comparing the concentration of chemicals in roots to the corresponding concentration in subsurface (Trapp, 2004). The most important factor to assess the uptake efficiency and membrane permeability in plants is transpiration stream concentration factor (TSCF), which is a measurement of a compound's ability to passively cross the root membrane (Dettenmaier et al., 2009). These two important factors are calculated by Equation (1) and Equation (2), respectively.

$$\mathrm{RCF} = \frac{C_{\mathrm{Shoot}},\ \text{chemical concentration in plant roots}}{C_{\mathrm{Media}},\ \text{chemical concentration in exposure medium}} \quad (1)$$

$$\mathrm{TSCF} = \frac{C_{\mathrm{Shoot}},\ \text{chemical concentration in plant shoot xylem}}{C_{\mathrm{Media}},\ \text{chemical concentration in exposure medium}} \quad (2)$$

In addition to these concentration factors, a variety of other concentration ratios have been also reported in literature, such as the fruit concentration factor (FCF), grass/soil accumulation factor (GSAF), leaf concentration factor (LCF), and stem concentration factor (SCF). Bioconcentration factors (BCFs) are used to relate chemical concentrations

measured in different plant tissues to concentrations in the soil supporting that plants or the exposure media. Translocation factor (TF) is used to measure the plants' potential for the translocation of chemicals from roots to shoots. Multiple research efforts to have been undertaken to better understand transmembrane permeability in plants by modeling TSCF and other important concentration factors (Dettenmaier, 2008; Pussemier, 1991). Most of these studies mainly focus on the physics of the problems, and have several limitations as models are plant- or compound-specific. The complexity of the soil-water-vapor nature of the root zone, and the countless combinations of organic compounds and plants result in an environmental system that is far complex for accurate modeling through a mechanistic approach. While plant and organic chemical interactions have been under intense study for almost years (Shone et al., 1974), new mechanistic models have been recently put forth (Dettenmeier, 2008). Thus, entirely new approaches to better elucidate this important environmental process are required, in particular as the water–food nexus is being globally impacted. Water scarcity has been increasing concurrently with food demands, and the shortening of the water cycle is increasing the potential for impacting our food supplies. Water reuse as irrigation waters is a prime example. Generating new TSCF and RCF values through uptake studies is a basic step for developing robust data sets for use in predictive models, and enhanced understanding over uptake of chemicals by plants.

1.3. ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

Predictive models have always been useful to address various problems in different fields of science and engineering. In particular, environmental models are needed for improving our understanding of current environmental cycles and processes, and then

predicting future impacts of global chances in water use, urbanization, climate change, and increasing anthropogenic footprints (new compound generation, increased wastage, and synergies of the exposome). Food production and water scarcity at a time of changes in our water, energy, and nutrient cycling are certainly among these complex systems that are in need of improved understanding.

Modeling and other related analyses can be also very useful to investigate the uptake and translocation of emerging contaminants in plants, again a very complex and dynamic system. Such analyses can be performed using novel intelligent techniques due to their advantages over traditional modeling approaches. Artificial intelligence (AI) revolves around the use of a set of unambiguous instructions that a computer can perform to solve a problem (Russell & Norvig, 2016). AI algorithms are powerful for learning from data since they enhance themselves by learning strategies that have worked well in the past. The solutions to a problem are obtained through a sequence of steps often determined by a trial-and-error exploration of alternatives using search algorithms (Korf, 1996). Machine learning (ML) is an application of AI, which makes a given system able to automatically learn and improve from experience without being explicitly programmed (Mitchell et al., 2013). ML requires searching through data to look for patterns and then it adjusts program actions accordingly.

To improve understanding of uptake and translocation of emerging contaminants in plants various AI and ML techniques are used in our studies. Neural networks (NN) are AI models constructed by a set of connected computational neurons so that each connection between neurons transmits a signal from one to another. Each computational neuron receives input information, processes the inputs and then send the output to all connected

neurons (Haykin et al., 2009). Fuzzy logic is another AI technique, which has four main components: fuzzification, fuzzy rules, fuzzy inference, and defuzzification (Zadeh, 1988). In the first step, crisp inputs (real values) are transformed into fuzzy inputs using membership functions. Fuzzy rules are constructed to infer the output based on input variables (Tan & Chua, 2007). Fuzzy inference system is responsible for the evaluation of fuzzy rules and combination of the results of the individual rules using fuzzy operations. Finally, defuzzification of the fuzzy outputs is performed using membership functions to obtain final real outputs (Kaur & Kaur, 2012). Neural networks and fuzzy logic are two capable techniques to process metadata and come up with robust models by which important conclusions can be drawn in terms of understanding uptake of contaminants with emerging concern.

Clustering is an unsupervised machine learning technique, which divides data into groups such that objects in the same group are more similar (Guha & Mishra, 2016). Clustering algorithms have several applications in different fields to uncover undetected relationships in a complex data. Partitional clustering is one of the most popular grouping approaches (Wilson et al., 2002). In partitional clustering, the data points are divided into a number of groups, and the goal is to minimize the sum of distances of all objects in the group from the center using similarity measures such Euclidean distance. Clustering algorithms can differentiate between plants with different uptake capacities, this open the avenues for several important further analyses regarding uptake and translocation of emerging contaminants in plants.

1.4. IMPORTANCE OF PLANT CHEMICAL UPTAKE

Contaminants with emerging concerns have received much attention in recent years due to their increasing occurrence in surface waters (Calderón-Preciado et al., 2011). Emerging contaminants through the application of reclaimed wastewater for crop irrigation, and use of sewage sludge on agricultural soil can be transferred to the food chain directly by uptake into food crops or indirectly following uptake into forage crops (Eggen et al., 2013). Human long-term exposure to emerging contaminants has received attention in recent years with the detection of these chemicals in blood and breast milk (Kärrman et al., 2007) and better understanding of emerging pollutants like PFAS (Sharifan et al., 2021). Considering the possible health risks associated with some of the emerging contaminants, which are very toxic and carcinogenic at low concentrations, a deeper look at the uptake and translocation of emerging contaminants is necessary for better assessment of human health risks. In addition to the food security and human health, uptake and translocation of emerging contaminants in plants is also important for the remediation of environmental contaminants. Phytoremediation of emerging contaminants is a cheap and environmental friendly approach to remove contaminants from polluted sites However, several challenges need to be addressed for an effective removal of contaminants by plants. Choosing the right plant species for the removal of emerging contaminants from polluted sites is the first challenge for the phytoremediation. Potentially poplars and willows are two plant species, which frequently are used for the in-situ decontamination of polluted sites (Robinson et al., 2000). However, knowledge is lacking about the difference between various plant species in terms of uptake and translocation of emerging contaminants. Investigating variability in plants is a basic step to successfully use plants for the removal

of contaminants from polluted sites, while at the same time understanding the potential for trophic level transfer and potential accumulation. For choosing the right species the interactions between plants and surrounding environment is important. The toxic impact of chemical compounds on the plant is an example, which can affect the removal of contaminants (Shahid et al., 2017). Toxic effects of chemicals can lower transpiration rate and through lowering plant health criteria decrease the uptake and translocation of contaminants. The role degradation is also important since endophytic bacteria (non-pathogenic bacteria that occur naturally in plants) and rhizospheric bacteria (live on and near the roots of plants) can contribute to the biodegradation of the toxic organic compounds (McGuinness and Dowling, 2009). Endophytes are microorganisms ubiquitously associated with almost all plants that dwell within robust plant tissues by having a symbiotic association. Endophytes produce a wide range of compounds, which are useful for plant growth and protection to environmental conditions (Nair and Padmavathy, 2014).

1.5. GOALS AND OBJECTIVES OF THIS STUDY

The over-arching and long-term goal of this study is to develop robust and intelligent models, which will be useful in the risk assessment associated with consumption of crops contaminated by organic compounds. The intelligent models by machine learning use physicochemical properties of contaminants significant to uptake and translocation of emerging and fugitive organic compounds, with insight to the fundamentals of transmembrane migration. The belief is that these intelligent models will significantly improve the understanding of uptake and translocation of organic compounds in plants, as

well as membrane permeability as well. To support the long term goal of the study, three main objectives are determined for my work:

Objective 1: Analyze significance of different physicochemical properties for plant uptake. Various physicochemical properties such molecular weight, hydrogen bond donor, hydrogen bond acceptor, rotatable bonds, polar surface area will be examined and their importance will be compared to log K_{ow}. Various statistical analyses such as stepwise regression will be conducted to investigate the significance of physicochemical properties for TSCF and RCF. The most significant physicochemical properties will be further examined to see if their interaction has impact of the uptake and translocation of contaminants by plants. For such analysis, the body of knowledge generated across hundreds of studies over decades of global research must be condensed for common use and assessment. To perform needed analyses, other machine learning techniques such as fuzzy logic and clustering algorithms will be used to examine such complex interactions. Fuzzy logic is a capable technique by which it is possible to examine the simultaneous effect of several physicochemical properties on the TSCF and RCF. The clustering algorithms are useful techniques to group various plant species based on their capabilities in the uptake and translocation of emerging contaminants.

Objective 2: Develop intelligent models to predict contaminants at various parts of the plants. The model will be based on artificial neural network (ANN) as a capable machine learning technique. The ANN will use significant compound properties and available metadata to construct predictive model. A part of the all data will be used to train ANN, a part for validation, and another for testing the predictive model. Despite high abilities of machine learning techniques in many fields, their application for the modeling

of uptake and translocation is absolutely lacking. The ANN model will significantly enhance the accuracy of predicting concentration factors such as TSCF, RCF and FCF. Such a predictive model is a useful tool for researchers globally to assess the risks associated with the long-term human exposure to the emerging contaminants in contaminated crops.

Objective 3: Generate new uptake data to fill specific chemical space that is unique. Following the guidelines for the uptake and translocation studies published by Environmental Protection Agency of the United States, new uptake data will be generated. The uptake studies are conducted by considering different plant species and chemical compounds with a wide range of physicochemical properties to generate a solid database. Due to the high cost of uptake studies and inconsistent data collection approaches, there are limited number of experimental TSCF values for some chemicals and plant species. By considering different plant species (tomato, corn, and wheat) and also chemicals from different groups with a wide range of physicochemical properties, the generated data will increase the data pool for future analyses. To improve the future models for uptake and translocation of emerging contaminants having a robust dataset, which is both consistent and accurate is fundamental. The generated data also will be useful for other analyses since the selected compounds are important due to their effect on human health.

Objective 4: Test and improve developed predictive models for the chemicals with future concerns. The available models mostly are compound-specific and also plant-specific, thus, they cannot be used for the prediction of a wide variety of chemicals. In addition, the low accuracy of predictive model is an obstacle for applying to untested compounds. The uptake models will be improved using the data obtained through the

uptake experiments. The predictive model will look at the chemicals that are new and novel, and provide potential for identifying future concerns, problems before they exist. The improved predictive models will be able to identify compounds that re outliers and plants that are outliers also. Such models with acceptable accuracies will be useful both for risk assessment and prescreening of new chemical compounds before conducting any experimental studies.

1.6. CORE HYPOTHESES

The central hypothesis of the work is that transmembrane migration and translocation of emerging contaminants are controlled by specific physicochemical properties of the compounds. Earlier mechanistic models consider various parameters for soil and plant species in addition to chemical properties. However, the accuracy of predictions has not increased significantly as compared to earlier more simple models. The Proposed hypothesis proposed that a poly parameter model that is relatively simple and easy to work with as compared to the highly complex mechanistic models requiring plant factors and properties as well. When the number of variables increases in the mechanistic models, the calibration of the models is more difficult and needs more detailed data and expertise.

Earlier studies consider lipophilicity (log K_{ow}) as the main property to control uptake and translocation of organic contaminants. While lipophilicity is important, the role of other physicochemical properties is thought to have specific impacts and should be considered. This hypothesis is supported by the fact that single parameter relationships based on lipophilicity cannot capture the complexity of the plant uptake, particularly when

dealing with unique solutes that are do not follow the predicted outcomes of the single parameter model. Considering other physicochemical properties can improve the accuracy of poly parameter predictive models for the uptake of emerging contaminants to support the core hypothesis.

The accuracy of predicting chemical uptake and translocation by intelligent models is hypothesized to significantly increase model prediction accuracy. The neural network models can consider several variables as their inputs while the network does not need calibration like mechanistic models. With the increased interactions between input variables and intelligent tuning algorithms, the accuracy of predictions should increase considerably and elucidate the importance of other chemical properties that contribute to controlling contaminant transmembrane migration in particular. The complex process of entry into and transport through membranes, has eluded clear understanding. These chemical properties are also thought to control accumulation in different plant compartments, such as roots, fruit, stem or leaves. Similar AI methods have been applied to problems in other fields of science and engineering, demonstrating neural networks models outperformed available mechanistic models of complex environmental systems, particularly for multimedia models with spatial and temporal variability. With improved accuracy of predicting concentration factors such as TSCF, RCF, and FCF, it is possible to assess uptake and translocation of emerging contaminants in plants. Fate of emerging contaminants in plants will have important implications to several problems including food security and human health risk assessments.

2. LITERATURE REVIEW

A vast number of plant uptake studies have been conducted to understand the fate of emerging contaminants in plants for roughly 50 years, with seminar work published in 1974 (Shone, 1974). Most of these studies have been conducted to introduce a relationship between the physical properties of organic chemicals and their translocation in plants. The early studies measured the uptake of various chemicals by different plant species. The uptake and subsequent translocation to shoots for two series of non-ionised chemicals (substituted phenylureas and O-methylcarbamoyloximes) in barely were found to be greater for more lipophilic chemicals (Briggs et al., 1982). These early reports characterized the behavior of many pesticides and organic pollutants in various plant species and presented derived empirical relationships and hypothesized mechanisms, however, the precise mechanisms were unknown. Many studies over decades presented various relationships of uptake and translocation when investigating different chemical classes and plants. The variability of the relationships indicated a single parameter model was not adequate to describe the uptake mechanistically.

The uptake of fungicides, herbicides and an insecticide by soybean roots, and their translocation through xylem were investigated using TSCF (Sicbaldi et al., 1997). By considering TSCF to assess efficiency of uptake and translocation, they observed a non-linear relationship between TSCF and log K_{ow}. The results for the uptake of four non-ionised compounds by aquatic plant parrot feather showed the importance of lipophilicity and underlined high uptake rates of the most lipophilic compounds (de Carvalho et al.,

2007). Translocation of the chemicals to shoots was optimal at log K_{ow} of 1.8, at which the efficiency of translocation of compound was about 40%.

With these experimental studies, different relationships were offered between the physicochemical properties and uptake efficiency. The results of these studies were contradictory in some cases. As an example, while some studies came up with a bell-shape relationship between log K_{ow} and TSCF (Briggs et al., 1982, Burken & Schnoor, 1998), the other reported a sigmoidal relationship between log K_{ow} and TSCF (Dettenmaier et al., 2009;). With such relationships, it was difficult to interpret the results for some compounds. For hydrophilic compounds (chemicals with log $K_{ow} < 0$), some models predicted high TSCF values while other models predicted low TSCF values. Moreover, there were several compounds as outliers, which single-parameter relationships based on log K_{ow} could not explain well.

Models developed for the fate and transport of contaminants in plants has followed two general approaches. In the earliest approaches, the TSCF is correlated with physicochemical properties of chemicals, mainly hydrophobicity (log K_{ow}). The other modeling approach is more complicated based on environmental compartments. Based on the simpler modeling approaches, the moderately hydrophobic compounds (chemicals with log K_{ow} of 1-3) have the greatest uptake efficiencies (Hsu et al., 1990). However, the more recent sigmoidal model offers the highest uptake efficiencies for the hydrophilic compounds in pressure chamber experiments (Dettenmaier et al., 2009). Recent investigations have incorporated the ionizable component of the compounds and compared uptake and translocation as a function of log D for previous studies (Schriever and Lamshoeft, 2020), including those mentioned above. The single parameter relationship was

again a bell shaped curve with improved correlation, yet in almost 50 years of investigation, the understanding of organic compounds had not notably improved over the initial empirical relationship. Uncertainty remains on the important aspect of organic pollutants entering into the transpiration stream and entering the base of the food chain.

These single-parameter relationships are oversimplified to provide the necessary insight into the fate and transport of emerging compounds in plants. Thus, to better explain the fate and translocation of emerging compounds in plants, additional chemical dimensions should be considered in the modeling approaches. Compartment-based models a more complicated approaches (overly parameter-intensive) are more flexible, and have data output and accuracy (Chiou et al., 2001; Collins and Finnegan, 2010; Trapp, 2007). The Compartment-based models rely on estimates of partitioning coefficients for the various compartments in which lipid content is a dominant factor (Chen et al., 2009; Zhang and Zhu, 2009). The transient nature of plant-contaminant system has complicated measurements of uptake efficiency and these modeling efforts that also must incorporate more plant-specific parameters (Rein et al., 2011).

In addition to the physicochemical properties, the role and importance of other factors related to soil properties and plant species are important. There have been several research efforts, which consider these factors in their uptake and translocation studies. The effect of soil pH on the movement and uptake of chlorsulfuron was investigated in different plant species (Fredrickson and Shea, 1986). According to their findings, at pH of 7.5 the growth of sorghum was not affected, while at pH of 5.9 there were large differences between treated plants and controls. The results also indicated that uptake of chlorsulfuron by wheat at pH of 5.9 is 67-100% greater as compared to pH of 7.5. The soil pH can

strongly affect the degradation rate of chlorsulfuron. In another study investigating the impact of soil water regime on plant uptake of the herbicide isoproturon, four different soil types were examined (Grundmann et al., 2011). The results indicated that the soil water regime is an essential factor that regulates degradation and plant uptake of isoproturon. It was also found that the intensity of this factor is strongly dependent on the soil type. In a more recent study, the effect of heavy metal concentrations in the soil was investigated on the concentrations of lead, cadmium and barium in urban garden-grown vegetables (McBride et al., 2014). The results demonstrated a poor relationship between vegetable and soil metal concentrations. This was attributable to particulate contamination of vegetables and soil characteristics, which influence phytoavailability.

The uptake and translocation of a given chemical compound also can be varied across various plant species. Partitioning to plant roots is a factor that affect the general uptake and translocation of emerging contaminants in plants. Plant root is considered to be constructed of water, wax, lignin, cellulose, lipid, phenolics, non-structural carbohydrates and ash as its fundamental materials (Martínez et al., 2002). The research studies have mainly investigated partitioning of various chemicals to lipid fraction of plant root as the driving factor. While lipids are important constituents for partitioning of chemicals into plant root tissues, the fraction of other constituents such as cellulose, lignin, and waxes are significantly high as compared to the lipids. Neglecting the role of these constituents is probably one of the main reasons why root-water partitioning relationships based on lipophilicity (log K_{ow}) underestimate the partitioning coefficients. The results of a recent study show that fraction of protein in the plant roots affects concentration factors including TSCF and RCF (Wen et al., 2016).

To develop predictive models for the uptake and translocation of emerging contaminants in plants, having a robust and consistent data set has always been one of the main obstacles (Fantke et al., 2016). Due to differences in the experimental design, methodology, and measurement of transpiration factors, the collected data for uptake and translocation studies is notably inconsistent, and numerous guidelines have been published to guide consistency in the collection of data (Doucette et al., 2018, EPA, 2012). The generated data is often less useful than it could be for direct use in the modeling purposes. In addition to the importance and role of driving factors for the assessment of plant uptake and translocation of emerging contaminants, lack of robust data set for developing plant bioaccumulation models is a serious issue, which need to be addressed appropriately (Gobas et al., 2016; McKone and Maddalena, 2007). However, through the modification of several existing guidelines, robust and comprehensive guidelines could be provided for the future uptake and translocation studies (Fantke et al., 2016). To achieve this basic goal, it is important to understand the measurements of vital parameters related to plant uptake and translocation of emerging contaminants that are required in bioaccumulation modeling.

Fate and transport of organic chemicals across biological membranes is an area of interest in mammalian systems. Based on the Lipinski's rule of five (Lipinski, 2004), the orally administered pharmaceuticals taken up by intestinal system are compounds with specific range of physicochemical properties. The rule of five states that compounds with 5 or fewer hydrogen bond donors, 10 or fewer hydrogen bond acceptors, a molecular mass of less than 500 Daltons, and log K_{ow} of less than 5 have the highest uptake capacities (Limmer and Burken, 2014). Such rules have been widely implemented in the drug development industry as prescreening filters to accelerate the drug development process

(Sicbaldi et al., 1997). Developing similar relationships for the uptake of emerging contaminants by plants has important implications to application of plants in drug development and other related problems. This can be supported by the fact that the physicochemical cutoffs for compound crossing plant root membrane were similar to physicochemical cutoffs offered by rule of five (Limmer and Burken, 2014). The similarity in uptake and translocation of organic compounds between plant roots and the human intestine has been shown in a previous study (Dettenmaier and Doucette, 2007). The findings of these studies and related research provide the foundation for importance of comparing the uptake of emerging contaminants by plants and uptake of orally administered drugs by intestinal system.

PAPER

I. A DEEPER LOOK AT PLANT UPTAKE OF ENVIRONMENTAL CONTAMINANTS USING INTELLIGENT APPROACHES

Majid Bagheri, Khalid Al-jabery, Donald Wunsch, and Joel G. Burken

Civil, Architectural and Environmental Engineering Department, Missouri University of Science and Technology, Rolla, MO 65409

ABSTRACT

Uptake of contaminants from the groundwater is one pathway of interest, and efforts have been made to relate root exposure to transloation throughout the plant, termed the transpiration stream concentration factor (TSCF). This work utilized machine learning techniques and statistcal analysis to improve the understanding of plant uptake and translocation of emerging contaminants. Neural network (NN) was used to develop a reliable model for predicting TSCF using physicochemical properties of compounds. Fuzzy logic was as a technique to examine the simultaneous impact of properties on TSCF, and interactions between compound properties. The significant and effective compound properties were determined using stepwise and forward regression as two widely used statiscal techniques. Clustering was used for detecting the hidden structures in the plant uptake data set. The NN predicted the TSCF with improved accuracy compared to mechanistic models. New insight was also delevered to compound properteis and their importance in transmembrane migration. The sensitivity anlysis indicated that log K_{ow}, molecular weight, hydrogen bond donor, and rotatable bonds are the most important

properties. The results of fuzzy logic demonstrated that the relationship between molecular weight and log K_{ow} with TSCF are both bell-shape and sigmoidal. The employed clustering algorithms all discovered two major distinct clusters in the data set.

1. INTRODUCTION

While many waters undergo advanced treatment to destroy recalcitrant organics, many other municipal wastewaters and agricultural run-off still contain considerable synthetic organic molecules (Hayashi et al., 2010; Hong et al., 2001; Tanoue et al., 2012). Irrigation of crops with reclaimed wastewater and application of agricultural chemicals have in part ameliorated water shortage problems and also enhanced agricultural productivity (Paraíba et al., 2010). Along with the irrigation of reclaimed wastewater and application of agricultural chemicals the exposure of plants to these compounds has increased (Aitchison et al., 2000; Dettenmaier and Doucette, 2007; Yoon et al., 2002).

The chemical contaminants in the soil have the potential of being transported to foliage through plant transpiration (Ciucani et al., 2002; Doucette et al., 2005a; Doucette et al., 2005b; Hsu et al., 1990; Orchard et al., 2000). Thus, quantification and prediction of transmembrane migration and transport from soil to foliage has direct linkage to potential human health impacts. (Russell and Shorrocks, 1959) introduced transpiration stream concentration factor (TSCF) to show the possibility of transporting a given chemical to foliage. TSCF is the ratio of a chemical concentration in the xylem sap to the concentration of that chemical in the solution passively transferred to leaves (Crowdy and Jones, 1956; Davis, 1998; Geissbühler et al., 1963; Su and Liang, 2011).

The high cost of experimental studies and inconsistent data collection have resulted in generating few number of experimental TSCF values for a limited number of chemicals and plant species (Fantke et al., 2016). The variability of the reported TSCF values for a given chemical and plant species is large due to the lack of consistent testing guidelines and difficulty in measuring metabolism and volatilization losses during the experiments (Edwards et al., 1982; Garvin et al., 2015; Yifru and Nzengung, 2006). The estimation of TSCF values for new contaminants not only helps to have predictive tools on efficiency of a specific molecule to be translocated by plants but also helps researchers focus efforts on contaminants with likely translocation capacity. Since 1974 several studies have been conducted to introduce a relationship between the physical properties of organic chemicals and their translocation in plants (Briggs et al., 1982; Briggs et al., 1987; Dettenmaier et al., 2008; Shone et al., 1974). These studies introduced single-parameter relationships relating the TSCF to octanol/water partition coefficient (log K_{ow}) which is as a term to describe hydrophobicity.

Due to the low precision of single-parameter relationships and the limitation of these relationships for applying to a wide variety of contaminants and plants species (Pussemier, 1991), researchers came up with more complicated prediction models for plant uptake and translocation. The single-parameter relationships were outperformed with the advent of recent models correlating multiple compound properties to TSCF (Limmer and Burken, 2014; Miller et al., 2016). More complex approaches are compartmental models (Collins and Finnegan, 2010; Manzoni et al., 2011; Undeman et al., 2009), which consider more chemical and environmental properties, and also incorporate the complexity of uptake and translocation processes into their mechanistic relationships. These modeling

approaches still have limited accuracy in many cases in spite of improving our understanding of plant uptake and translocation of contaminants. However, in the majority of these modeling efforts, the models are calibrated or verified with specific plant species and chemicals tested in the laboratory portion. Plant selection and experimental design can certainly impact findings and modeling efforts. The current field of contaminant uptake has developed a large data pool (Felizeter et al., 2014; Limmer and Burken, 2014; Miller et al., 2016) that can be used to investigate the comprehensive data sets for uptake of a wide array of compounds, by a range of plants, and in multiple laboratory arrangements, thereby limiting impacts of any one arrangement or data set. The assessment of these large data agglomerations can be challenging given the complexity of the data, and thus needing advanced data assessment methods and tools.

Simulation has been a useful approach to deal with various problems in different fields of science and engineering (Sofizadeh et al., 2016; Trapp et al., 1994). In this work, a neural network (NN) predicts the plant uptake and translocation of environmental contaminants. The NN uses physicochemical properties of compounds to assess past data collections and predict TSCF. The physicochemical properties of compounds are analyzed using statistical analysis to determine the importance of each property to the TSCF values. Fuzzy logic was used to examine the interactions between important physicochemical properties as predictors of experimentally-determined TSCF values. Furthermore, clustering techniques were utilized to determine any distinct groups and hidden data structures in the comprehensive data set.

2. MATERIAL AND METHODS

2.1. PLANT UPTAKE DATA SET

There are considerable number of published studies, which have investigated the uptake of various compounds (Crowdy and Pramer, 1955; Farlane et al., 1990; Kim et al., 2004; Qiu et al., 2016a; Qiu et al., 2016b; San Miguel et al., 2013; Sheets, 1961; Thompson et al., 1999). In order to build the NN model, a comprehensive selection of TSCF data was compiled from published literature (Supporting Information). The selected TSCF data set to build general model includes 300 measurements of 155 compounds measured using 34 plant genera under various experimental approaches from 41 studies. The measurements for TSCF were not included when there was no evidence of reaching the steady state, measurements included metabolites, there was metabolism of the parent compound in planta, or the TSCF calculation was not reliable. Moreover, the measurements were not included when depletion of the dosing solution was higher than 50%, roots were damaged prior to dosing, or other modes of exposure such as particle deposition were included in the measurements (Limmer and Burken, 2014). To build a general model that is not plant specific, the compound properties including log K_{ow}, molecular weight (MW), hydrogen bond donor (HBD), hydrogen bond acceptor (HBA), rotatable bonds (RB) and polar surface area (PSA), were obtained from chemical structure databases such as US EPA Chemistry Dashboard, and ChemSpider. For the properties like log K_{ow} that both predicted and experimental values are available, by considering the range of variation for the properties, the experimental values were considered in the analysis. Table 1 shows the

compound properties used as input parameters in this study for the neural network modeling and other analyses.

Table 1. Characteristics of measured parameters used in the NN modeling process.

Input variable	**Min-Max**	**Output variable**	**Min-Max**
Log K_{ow}	-2.19–6.75	**TSCF**	0.001-1.16
MW (Da)	32–616.4		
HBD	0–6		
HBA	0–16		
RB	0–36		
PSA (A^2)	0–196.2		

2.2. NEURAL NETWORK MODEL

A NN is a massively parallel distributed processor consisting of simple processing units that have the ability to learn from experience (Haykin et al., 2009; Samarasinghe, 2016). The NN implemented in this work consists of three layers of interconnected neurons. A single-output NN with M neurons in the hidden layer is expressed by Equation (1):

$$y(w,x) = \varphi_{out}(\sum_{i=1}^{M}(W_{i,out} \times x_i) + b_{out}) \quad (1)$$

where, φ_{out} is the activation function of the output layer, $W_{i,out}$ is the weight between the ith neuron in the hidden layer and the output neuron, b_{out} is the bias of output neuron, and x_i is the output of each neuron in the hidden layer and is calculated by Equation (2):

$$y_m = \varphi_h(\sum_{i=1}^{M}(W_{i,m} \times x_i) + b_m) \quad (2)$$

where, φ_h is the activation function of hidden layer, M is the number of input parameters, $W_{i,m}$ is the weight between the ith input parameter and the mth neuron in the hidden layer, b_m is the bias of mth neuron in the hidden layer, and x_i is the ith input parameter.

In this study, a NN was used for the first time to predict the uptake and translocation of emerging contaminants in plants. The NN modeling was utilized due to its higher accuracy than current mechanistic models (Manzoni et al., 2011). The input layer of the NN model consisted of six neurons for each of the six inputs (log K_{ow}, MW, HBD, HBA, RB, and PSA). These molecular descriptors were considered as inputs of the NN because the properties have been cited as important parameters in the modeling of TSCF in recent studies (Limmer and Burken, 2014; Miller et al., 2016). The network architecture that was implemented in this work is a feedforward NN. The dataset was divided randomly into three parts, 70% for training, 15% for testing, and 15% for validation of the NN model. The code was written in MATLAB R2014a. More information regarding various NNs can be found in (Haykin et al., 2009). The prediction performance of the NN model for the TSCF was measured using correlation coefficient (R) and mean squared error (MSE).

2.3. STATISTICAL ANALYSIS

In this study, stepwise and forward regression were used to determine the most important predictors (compound properties) in modeling TSCF. Forward regression is the simplest data-driven selection approach in which one predictor is added to the model at a time. Forward regression starts with no predictors in the model and calculates the p-value (a criterion for selection) for all predictors not in the model (Guo et al., 2015). It adds the predictors with p-value less than p-critical (5% as a widely accepted value), and this

process is repeated until no new variable can be selected and added to the model. Stepwise regression is a modification of forward regression. In this method, a predictor may be added or removed from the model at a time. Like forward regression, it starts with no predictors in the model and calculates the p-value for all predictors not in the model. After adding a predictor to the model, all variables in the model are examined to find if their importance has been reduced to a predefined limit (p-critical of 5% in this study) of selection (Nazarpour et al., 2016). The predictors in the model with importance less than a predefined limit are removed from the model. Forward and stepwise regression were performed using SPSS 16.0 software in order to determine the significant predictors.

2.4. FUZZY LOGIC

Fuzzy logic is an intelligent technique based on degrees of truth rather than true or false logic. Fuzzy logic is distinguished from familiar approaches such as Boolean algebra due to its ability to present results in the form of recommendations (Zadeh, 1965). Fuzzy logic has been widely used for solving problems in environmental engineering and other fields as well (Bagheri et al., 2017). Fuzzy logic is conducted through four main steps: fuzzification, generating fuzzy rules, generating a fuzzy inference system, and defuzzification. The fuzzification is a process to transform crisp inputs into ordinary fuzzy inputs using membership functions. The non-fuzzy input values are converted to fuzzy linguistic terms using triangular, trapezoidal, piecewise linear, Gaussian, and other functions (Klir and Yuan, 1995). The type of the membership function for each problem is chosen arbitrarily based on the experience of users. Fuzzy rules (simple IF-THEN rules with a condition and a conclusion) are constructed to infer the output variable based on

input variables. The inputs to a simple IF-THEN rule are the current values for the input variables and the output is an entire fuzzy set. The fuzzy rules for a system with two inputs of X1 and X2, and one output of Y, can be expressed by Equation (3):

$$IF \quad X_1 \; is \; A \quad and \quad X_2 \; is \; B \quad THEN \quad Y = aX_1 + bX_2 + c \tag{3}$$

where, A and B are linguistic labels, a, b and c are constants that are adjusted by optimization algorithms.

After evaluation of the fuzzy rules, the final result is obtained by combining the results of fuzzy rules through a process called inference (Von Altrock, 1995). Takagi-Sugeno and Mamdani are two widely used inference systems. The main difference between these two inference systems is the way crisp output is generated from the fuzzy inputs. In the last step, the results obtained in the form of fuzzy sets are converted into numerical values using membership functions. More discussion on fuzzy logic can be found in (Klir and Yuan, 1995; Kosko, 1997; Ross, 2009).

In this study, fuzzy logic was used to project a general relationship between compound properties and TSCF with the aid of available plant uptake data. Fuzzy logic also was utilized as an appropriate technique to examine the simultaneous impact of properties on TSCF. Fuzzy logic helps to screen the capacity of chemical compounds for uptake and translocation in plants through correlating TSCF values with various compound properties at the same time. An adaptive neuro-fuzzy inference system (a Takagi-Sugeno fuzzy rule base), and various membership functions were employed to develop fuzzy logic using MATLAB R2014a.

2.5. CLUSTERING ALGORITHMS

Clustering algorithms are capable tools to reveal hidden relationships and structures in data (Xu and Wunsch, 2009). In this work, two types of clustering algorithms including hierarchical and partitional clustering were applied to detect the hidden relationships in plant uptake data. Hierarchical clustering results in a set of nested clusters that are organized as a tree, and the generated tree may correspond to a meaningful taxonomy. Through partitional clustering, data objects in the data set are divided into non-overlapping subsets such that each data object is in exactly one subset. The applied partitional clustering algorithm is the well-known k-means that divides data objects in the plant uptake data set into k distinct clusters. Further discussion on these approaches can be found in (Xu and Wunsch, 2009). Internal validation indices and visualization-based methods were used to compare between the clustering algorithms, and also to obtain the best clusters. The applied internal validation indices are Silhouette, Davies-Bouldin, and Calinski-Harabasz. The mathematical formulas and further discussion on the above cluster validation indices can be found in (Liu et al., 2010; Xu and Wunsch, 2005; Xu and Wunsch, 2009). The resulting clusters in our study are achieved by considering the six physicochemical properties of the compounds. In order to provide a descriptive visualization, principal component analysis (PCA) was used to visualize the resulting clusters from different algorithms using the first two components. Through PCA visualization, data is transformed from higher dimensions to lower dimensions (two dimensions in this study) without losing a significant amount of information (Wold et al., 1987).

3. RESULTS AND DISCUSSION

3.1. PREDICTION OF TSCF USING NN

The NN learned an acceptable fit between predicted and measured values, as the regression line for measured values of TSCF and predicted values of TSCF were in agreement (Figure 1). The correlation levels between predicted and measured values of TSCF were approximately 0.83, 0.73 and 0.79 for training, validation and testing data sets, respectively, as illustrated in Figure 1. The minimum calculated MSE was 0.016, 0.059, and 0.037 for training, validation, and testing data sets, respectively. These results indicate that the NN has adapted well during training and can be used as a general model for predicting TSCF using the given parameters, whereas traditional models were proven to be inaccurate for some compounds (Miller et al., 2016). For example, (Fujisawa et al., 2002) developed uptake models for translocation of neural pesticides to the stem and edible parts of the crops. They validated the models using six pesticides with soybean and spinach plants. While these models were plant and compound specific, the ratio of the predicted to the measured concentration was 0.44-1.49 and 0.57-2.93 with foliage and seed models, respectively. It should be noted that the training procedures of NN models is based on the training data set. The NN does not have access to the test data set in this step and this part of the data is used to measure the accuracy of predictions. To check the accuracy of NN models before the final and optimized network, the validation data set is used. The validation data set is a part of data that NN uses to check if the training procedure is on the right track or not. Thus, the good correlation between predicted and measured TSCF based on test data shows the reliability of the developed NN model.

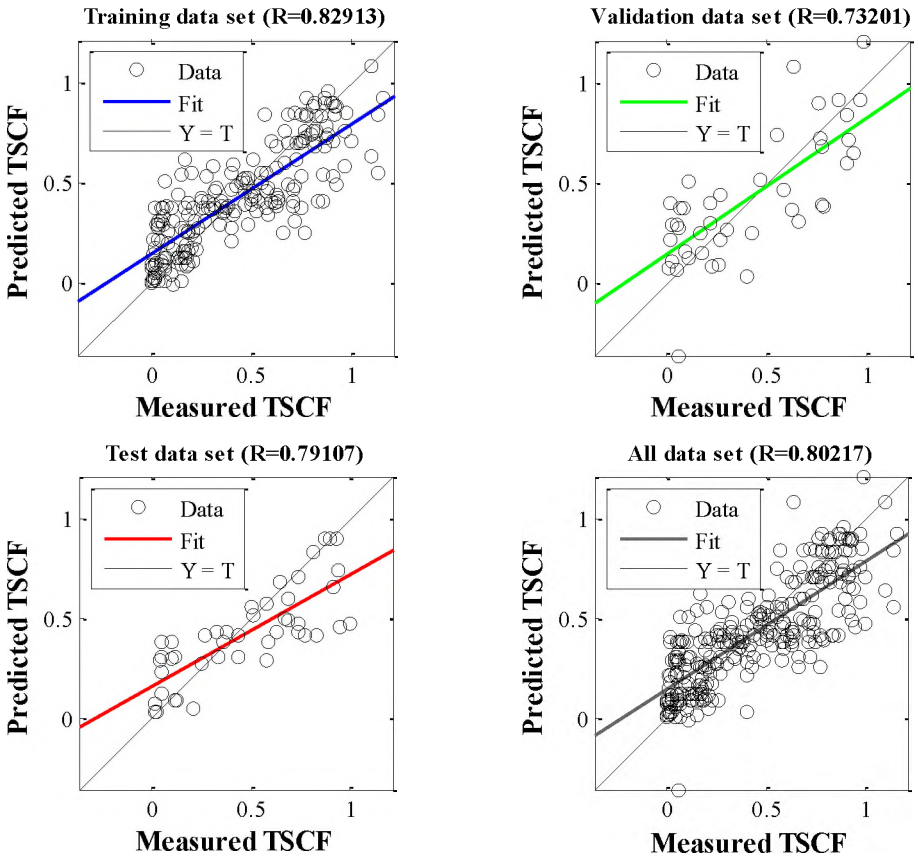

Figure 1. Regression plots of the NN model for training, validation, testing, and all data sets.

The normal distribution of error reveals that the predictions by the NN model were symmetrical and their axis approaching zero (Figure 2). The determination of a normal distribution of residuals is important, as such distribution is one of the assumptions for regression analysis (Montgomery et al., 2012), and validates the lack of bias in NN model predictions. The NN is able to predict the plant uptake and translocation of environmental contaminants with higher accuracy than traditional modeling approaches (Miller et al., 2016). The testing model demonstrated a positive prediction performance in comparison with traditional models, and the MSE value for this NN model was 0.037 vs. 0.25 to 0.3

for traditional models (Limmer and Burken, 2014; Miller et al., 2016). The NN learned to predict the TSCF from the six compound properties, and captured the general changes in TSCF values with no notable oscillations, as shown in Figure 3. Earlier modeling approaches reliant on a single parameter have shown inconsistency in predicting TSCF values, notably for hydrophilic compounds (Miller et al., 2016). The results of previous studies show that the models did not capture the general changes of TSCF.

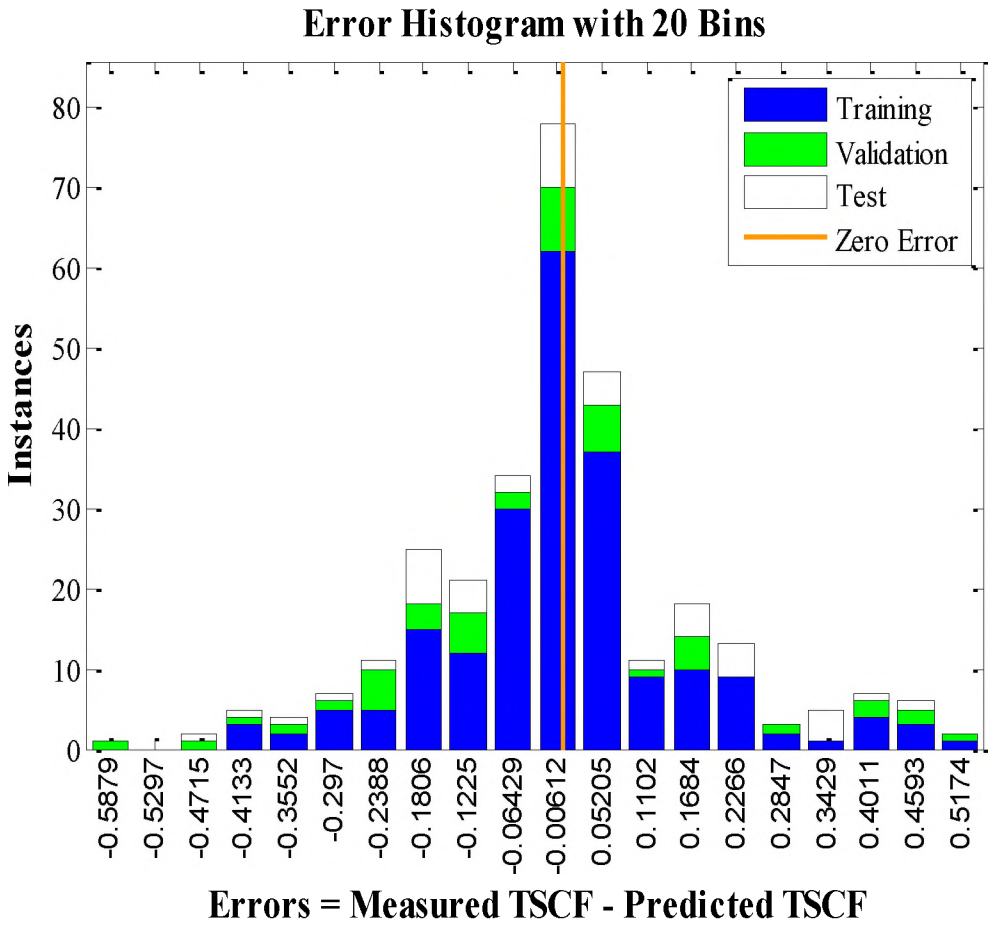

Figure 2. Residual of the NN model for training, testing, and validation data sets.

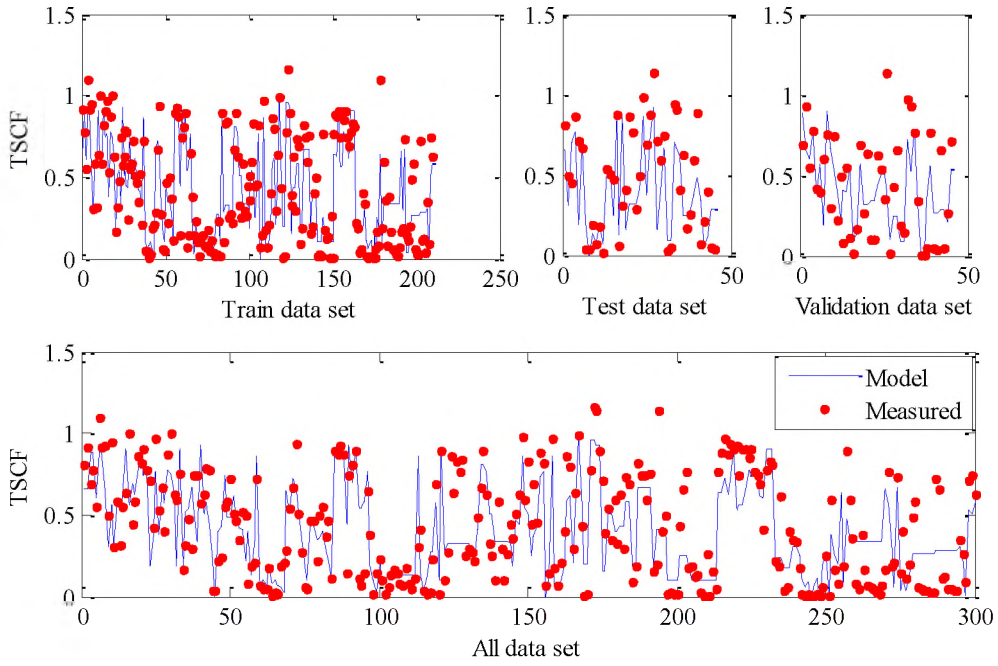

Figure 3. Predicted TSCF values by the NN model for training, validation, testing, and all data sets.

3.2. SIGNIFICANT COMPOUND PROPERTIES

Stepwise and forward regression methods were used to analyze the sensitivity of TSCF to the input parameters including log K_{ow}, MW, HBD, HBA, RB, and PSA. The results of sensitivity analysis were highly correlated for both methods. Both selection methods confirmed that log K_{ow} is the most important compound property (predictor). With a p-value less than 0.05 and t-statistics equal to 6.36 and 1.86, the log K_{ow} and MW are two compound properties with highly significant impacts for the plant uptake and translocation models. The HBD was the third important compound property (predictor) for the plant uptake and translocation. The t-statistics value for HBD was 1.61, as illustrated in Table 2. The results of this study were in line with the findings of previous studies trying to

determine the importance of physicochemical predictors. Using a desirability function indicated that log K_{ow}, MW, and HBD are the three important predictors in the modeling of plant uptake and translocation of organic contaminants (Limmer and Burken, 2014). Based on the results of our regression methods, the log K_{ow} is eight times more governing than MW as a predictor. The MW is also three times more important than HBD. The results of the stepwise and forward regression methods show that the RB is the fourth most effective predictor. In a recently published paper, Miller et al. (2016) used a desirability model for plant uptake of pharmaceutical and personal care products. They found that the number of rotatable bonds is another important predictor. The findings of this study indicate that RB is a predictor with almost the same importance as HBD in the modeling of plant uptake and translocation of contaminants. The correlation between these significant and important properties has been presented in Table 2.

Table 2. Results of sensitivity analysis for the significant and important compound properties.

		log K_{ow}	**MW**	**HBD**	**RB**
Stepwise	P-value	0.001	0.044	0.107	0.258
	t-statistics	6.368	1.861	1.617	1.133
Forward	P-value	0.001	0.045	0.109	0.258
	t-statistics	6.364	1.855	1.614	1.133
Correlation	**log K_{ow}**	1	0.432	-0.322	0.131
	MW	0.432	1	0.237	0.595
	HBD	-0.322	0.237	1	0.203
	RB	0.131	0.595	0.203	1

3.3. SIMULTANEOUS IMPACTS OF COMPOUND PROPERTIES USING FUZZY LOGIC

The results of fuzzy logic for examining the interaction between compound properties improved our understanding of plant uptake and translocation problems. Previous studies introduced a bell-shape relationship between log K_{ow} and TSCF (Limmer and Burken, 2014). There are a number of studies that found various optima for the most translocatable compounds by focusing on the log K_{ow}. Behrendt and Brüggemann (1993) reported that chemical uptake has a pronounced optimum for moderately lipophilic chemicals with log K_{ow} equals to 1.78. Burken and Schnoor (1998) also found that TSCF for organic compounds is maximum at log K_{ow} of 1.8. Sicbaldi et al. (1997) reported that for the pesticides from different chemical classes, the highest TSCF values are achieved for the compounds with log K_{ow} value around 3. Based on the review of conducted studies, Dettenmaier et al. (2008) concluded that that either highly polar compounds (log $K_{ow} < 1$) or highly lipophilic compounds (log $K_{ow} > 4$) are not significantly taken up by plants. The results of our study while are supported by the findings of previous studies, improved the understanding of plant uptake and translocation of contaminants. The most readily translocated compounds were relatively hydrophobic compounds with log K_{ow} between 1 and 4. The MW and TSCF had a bell-shape relationship, and the MW was less 350 Da for the most trans-locatable compounds. Figure 4 demonstrates the relationship between both log K_{ow} and MW with TSCF using an adaptive neuro-fuzzy inference system. The results of our study also confirm that generally there are bell-shape relationships between both log K_{ow} and MW with TSCF. However, the interaction between log K_{ow} and MW is important and affects the translocation of compounds.

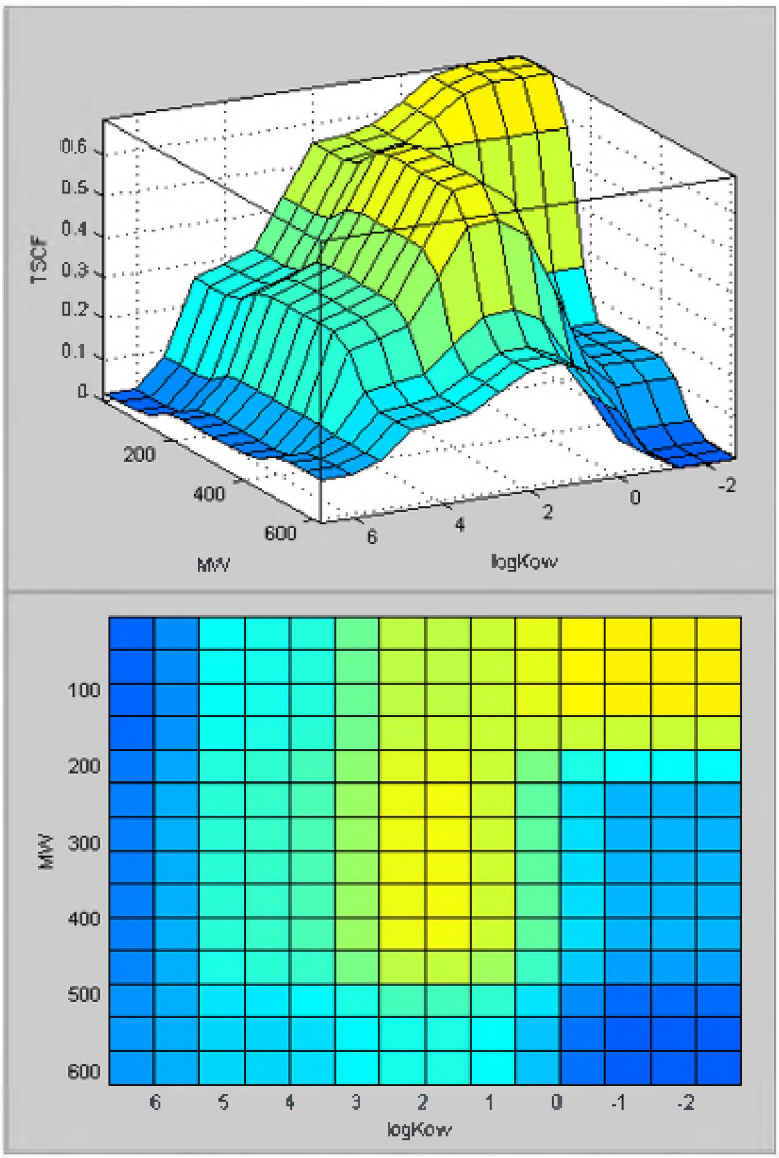

Figure 4. Relationship between log Kow and MW with TSCF using adaptive neuro-fuzzy inference system.

In addition to the log K_{ow}, MW is another important predictor that have received attention in a number of studies. Hong et al. (2001) reported that low MW hydrophilic compounds such as methyl tert-butyl ether transport through plant membranes via hydrogen bonding with water or by rendering plant membranes more permeable. Boonsaner et al. (2011) also confirmed the importance of MW and reported that since it

reflects the molecular size of a compound can influence the absorption process through the cell membrane of the plants. They found that for the benzene, toluene, ethylbenzene and xylenes with MW of 78 Da to 106 Da, the TSCF is high and increases with the MW. These studies agreed with (Connell, 1990) who reported increase in the bioaccumulation of organic chemicals with the increase of MW, but decline only for chemicals with MW of 350 Da and higher. In this study, it was observed that for the MW less than 120 Da, the relationship between log K_{ow} and TSCF is sigmoidal instead of bell-shape (Figure 4). For the MW from 150 to 400 Da, the relationship between log K_{ow} and TSCF is bell-shape with high possibility of translocation. For the MW higher than 400 Da, the relationship between log K_{ow} and TSCF is still bell-shape with lower possibility of translocation for compounds. The results of this study indicated that for the log K_{ow} less than 1, the relation between MW and TSCF is sigmoidal instead of bell-shape. The compounds have a high tendency for translocation in plants when the MW is less than 120 Da and log K_{ow} less than 1. These findings are interesting since previous studies have reported the translocation of hydrophilic compounds in addition to the moderately hydrophilic compounds (Dettenmaier et al., 2009; Trapp, 2007). Moreover, these findings are useful for assessing human health risks associated with emerging contaminants, food safety, and efficient phytoremediation of environmental contaminants (Boonsaner et al., 2011; Krstich and Schwarz, 1990; Rubin and Ramaswami, 2001; Su et al., 2010).

The findings of previous studies using the desirability function show a sigmoidal relationship between the HBD and TSCF. The desirability function demonstrated that compounds with HBD equal or less than 5 have a higher possibility for uptake and translocation (Limmer and Burken, 2014; Miller et al., 2016). Figure 5 demonstrates the

relationship between both log K_{ow} and HBD with TSCF using an adaptive neuro-fuzzy inference system. In a good agreement with the results of previous studies, the log K_{ow} has a bell-shape, and HBD has a sigmoidal relationship with TSCF. The results show that compounds with 5 or less HBD have a high potential for uptake and translocation. The compounds with 1 or 2 HBD have highest capacity for uptake and translocation in plants.

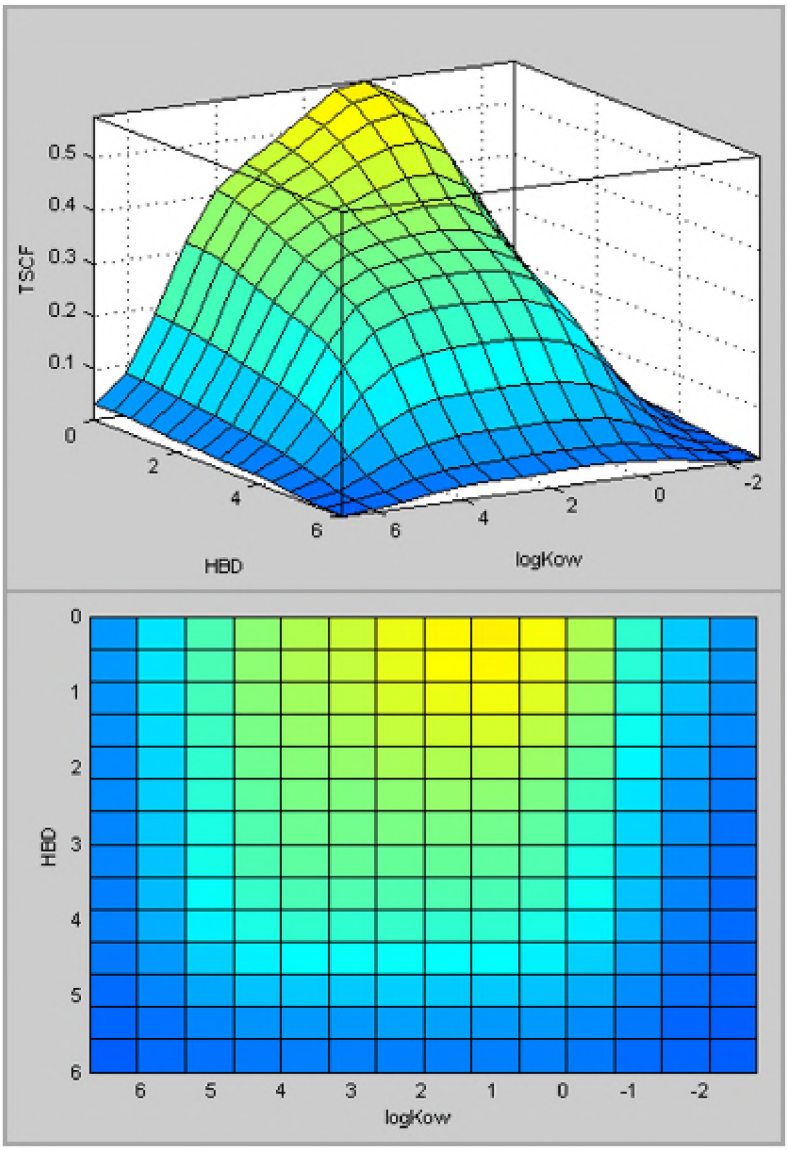

Figure 5. Relationship between log Kow and HBD with TSCF using neuro-fuzzy inference system.

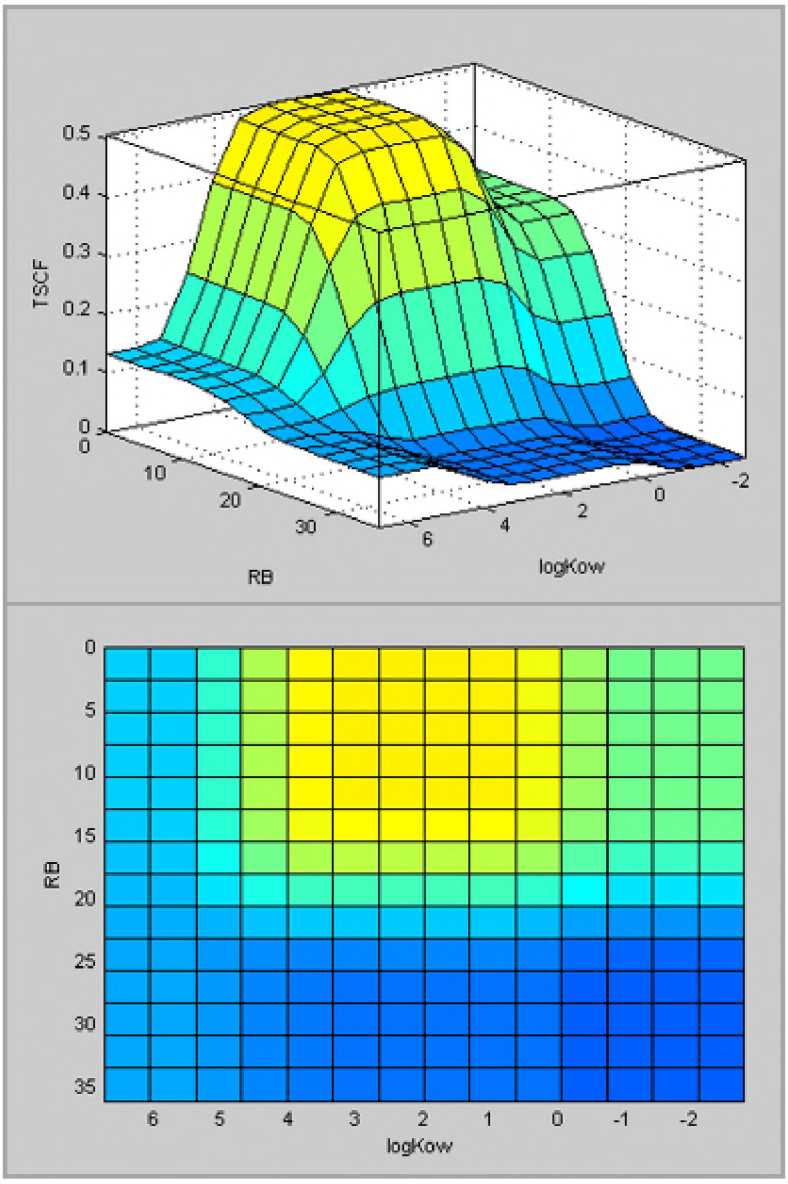

Figure 6. Relationship between log Kow and RB with TSCF using neuro-fuzzy inference system.

The desirability function also demonstrated that there is a sigmoidal relationship between rotatable bonds and TSCF. The desirability function showed that compounds with 7 or less RB are more likely to be translocated in plants. The results of our study using fuzzy logic conform the sigmoidal relationship between RB and TSCF (Figure 6). It was

observed that compounds with 15 or less RB are more likely to be translocated by plants. Moreover, for this range of RB, the compounds with log K_{ow} less than zero have a higher possibility of translocation than compounds with log K_{ow} higher than 4. The value of RB for high possibility of translocation from this study is a little different from the findings of desirability functions. Although our database is the most comprehensive database ever used for modeling of plant uptake and translocation of environmental contaminants, there are limited number of compounds with RB higher than 15. Thus, with increase in the number of published data for all chemical compounds, the accuracy of modeling approaches increases accordingly.

3.4. RESULTING CLUSTERS

Since the compiled dataset includes a considerable number of various compounds and plant species, clustering algorithms were used to examine the chemical compounds and plants that have more distinct behavior than others. All the applied algorithms have isolated six samples sharing almost no variables with the remaining data objects in the plant uptake data set, see *cluster 3* in Figure 7. Based on the statistical metrics, the k-means algorithm clustered the data objects into three distinct groups with different features. The results show that *cluster 3* was distinct from *cluster 1* and *cluster 2* in terms of chemical compounds and physicochemical properties of the compounds. The compounds in *cluster 3* have log K_{ow} values higher than 4, and an RB value of more than 21 to 32. The compounds in *cluster 3* have shown a low capacity for uptake and translocation with TSCF values mainly less than 0.05. Such compounds have a small capacity to cross the membrane of the plant roots and mainly accumulate in the root (Burken and Schnoor, 1998).

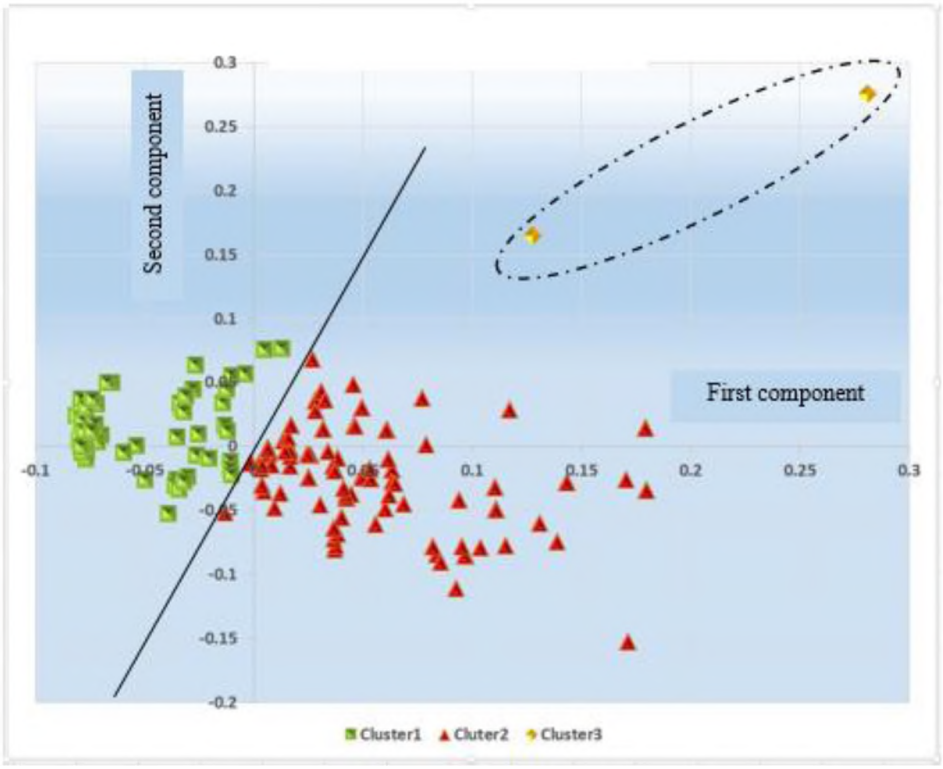

Figure 7. Obtained clusters by k-means algorithm (k=3) after visualization of results using PCA.

The clusters analysis also showed interesting results for tomato, wheat, and corn as three plant species, which are used in many uptake experimental studies. Based on different thresholds of uptake considered in this study (TSCF from 0.4 to 0.9), the tomato and wheat mostly were in the group with higher uptake and translocation capacity than the considered limit. The corn almost in all cases was in the group with less uptake and translocation capacity than the considered limit. Since corn has lower transpiration and higher growth, this finding is in agreement with theoretical concept for TSCF outlined in (Trapp, 2007), which predicts higher TSCF for less growth per transpiration. Such a finding using clustering algorithms is interesting since clustering does not care about the physic of problem and detect the hidden relationships just based on uptake data, and considered

variables. To have even more meaningful clusters regarding uptake and translocation of contaminants considering plant properties in addition to physicochemical properties of compounds is recommended in future research studies.

4. CONCLUSIONS

The introduction of mathematical relationships relating compound properties to the possibility of compounds uptake and translocation in plants has received considerable attention since 1974. Most of previous studies with the focus on compound properties introduced one dimensional relationships to relate log K_{ow} to TSCF. However, these relationships are limited to specific compounds and plant species, and are inaccurate in some cases. To the best of our knowledge, this study is the first utilizing machine learning algorithms (i.e. NN, Fuzzy logic and unsupervised learning) to predict TSCF and examine the interaction between compound properties. The NN predicted the TSCF with higher accuracy than previous models using physicochemical properties of compounds, and also offered insight to the impact and interactions of the different parameters. The log K_{ow}, MW, HBD, and RB were the most important compound properties in governing uptake of organic molecules. The findings indicate that log K_{ow} is the most significant property, MW is the second most important, and HBD and RB are two properties with similar level of importance. The results of fuzzy logic indicated that interaction between compound properties is an important factor to consider. The log K_{ow} and MW generally had bell-shape and sigmoidoid (little bell-shape and mostly sigmoidal) relationships with TSCF, not just a bell-shape or a sigmoidal one. Clustering algorithms also revealed previously

undiscovered structures in the dataset. The findings of this study shows the importance of hydrophilic compounds in addition to the moderately hydrophobic compounds. This study has also provided an estimation based on statistical evaluation metrics for TSCF threshold.

REFERENCES

Aitchison, E.W., Kelley, S.L., Alvarez, P.J., Schnoor, J.L. (2000). Phytoremediation of 1, 4-dioxane by hybrid poplar trees. *Water Environment Research*, 72, 313-321.

Bagheri, M,. Bazvand, A., Ehteshami, M. (2017). Application of artificial intelligence for the management of landfill leachate penetration into groundwater, and assessment of its environmental impacts. *Journal of Cleaner Production*, 149, 784-796.

Behrendt, H., Brüggemann, R. (1993). Modelling the fate of organic chemicals in the soil plant environment: model study of root uptake of pesticides. *Chemosphere*, 27, 2325-2332.

Boonsaner, M., Borrirukwisitsak, S., Boonsaner, A. (2011). Phytoremediation of BTEX contaminated soil by Canna generalis. *Ecotoxicology and Environmental Safety*, 74, 1700-1707.

Briggs, G.G., Bromilow, R.H., Evans, A.A. (1982). Relationships between lipophilicity and root uptake and translocation of non-ionised chemicals by barley. *Pest Management Science*, 13, 495-504.

Briggs, G.G., Rigitano, R.L., Bromilow, R.H. (1987). Physico-chemical factors affecting uptake by roots and translocation to shoots of weak acids in barley. *Pest Management Science*, 19, 101-112.

Burken, J.G., Schnoor, J.L. (1998). Predictive relationships for uptake of organic contaminants by hybrid poplar trees. *Environmental Science and Technology*, 32, 3379-3385.

Ciucani, G., Trevisan, M., Sacchi, G.A., Trapp, S.A.J. (2002). Measurement of xylem translocation of weak electrolytes with the pressure chamber technique. *Pest Management Science*, 58, 467-473.

Collins, C.D., Finnegan, E. (2010). Modeling the plant uptake of organic chemicals, including the soil– air– plant pathway. *Environmental Science and Technology*, 44, 998-1003.

Connell, D. (1990). *In Bioaccumulation of Xenobiotic Compounds*. CRC Press, Boca Raton.

Crowdy, S., Pramer, D. (1955). The Occurrence of Translocated Antibiotics in Expressed, Plant Sap. *Annals of Botany*, 19, 79-86.

Crowdy, S., Jones, D.R. (1956). The translocation of sulphonamides in higher plants: I. Uptake and translocation in broad beans. *Journal of Experimental Botany*, 7, 335-346.

Davis, L.C.V. S., Dana, J., Selk, K., Smith, K., Goplen, B., Erickson, L. E. (1998). Movement of Chlorinated Solvents and Other Volatile Organics through Plants Monitored by Fourier Transform Infrared (FT-IR) Spectrometry. *Journal of Hazardous Substance Research*, 1, 1-26.

Dettenmaier, E., Doucette, W.J. (2007). Mineralization and plant uptake of 14C-labeled nonylphenol, nonylphenol tetraethoxylate, and nonylphenol nonylethoxylate in biosolids/soil systems planted with crested wheatgrass. *Environmental Toxicology and Chemistry*, 26, 193-200.

Dettenmaier, E.M., Doucette, W.J., Bugbee, B. (2008). Chemical hydrophobicity and uptake by plant roots. *Environmental Science and Technology*, 43, 324-329.

Dettenmaier, E., Doucette, W., Bugbee, B. (2009). Chemical hydrophobicity and uptake by plant roots. *Environmental Science and Technology*, 43, 324-329.

Doucette, W.J., Wheeler, B.R., Chard, .JK., Bugbee, B., Naylor, C.G., Carbone, J.P., Sims, C.R. (2005). Uptake of nonylphenol and nonylphenol ethoxylates by crested wheatgrass. *Environmental Toxicology and Chemistry*, 24, 2965-2972.

Doucette, W.J., Chard, J.K., Moore, B.J., Staudt, W.J., Headley, J.V. (2005). Uptake of sulfolane and diisopropanolamine (DIPA) by cattails (Typha latifolia). *Microchemical Journal*, 81, 41-49.

Edwards, N., Ross-Todd, B., Garver, E. (1982). Uptake and metabolism of 14C anthracene by soybean (Glycine max). *Environmental and Experimental Botany*, 22, 349-357.

Fantke, P., Arnot, J.A., Doucette, W.J. (2016). Improving plant bioaccumulation science through consistent reporting of experimental data. *Journal of Environmental Management*, 181, 374-384.

Farlane, C.M., Pfleeger, T., Fletcher, J. (1990). Effect, uptake and disposition of nitrobenzene in several terrestrial plants. *Environmental Toxicology and Chemistry*, 9, 513-520.

Felizeter, S., McLachlan, M.S., De Voogt, P. (2014). Root uptake and translocation of perfluorinated alkyl acids by three hydroponically grown crops. *Journal of Agricultural and Food Chemistry*, 62, 3334-3342.

Fujisawa, T., Ichise, K., Fukushima, M., Katagi, T., Takimoto, Y. (2002). Improved uptake models of nonionized pesticides to foliage and seed of crops. *Journal of Agricultural and Food Chemistry*, 50, 532-537.

Garvin, N., Doucette, W.J., White, J.C. (2015). Investigating differences in the root to shoot transfer and xylem sap solubility of organic compounds between zucchini, squash and soybean using a pressure chamber method. *Chemosphere*, 130, 98-102.

Geissbühler, H., Haselbach, C., Aebi, H., Ebner, L. (1963). The fate of N'-(4-CHLOROPHENOXY)-PHENYL-NN-DIMETHYLUREA (C-1983) in soils and plants. *Weed Research*, 3: 277-297.

Guo, Y., Guo, L., Billings, S.A., Wei, H-L. (2015). An iterative orthogonal forward regression algorithm. *International Journal of Systems Science*, 46, 776-789.

Hayashi, O., Kameshiro, M., Satoh, K. (2010). Intrinsic bioavailability of 14C-heptachlor to several plant species. *Journal of Pest Science*, 35, 107-113.

Haykin, S.S. (2009). *Neural networks and learning machines* (Third Ed.). Pearson Education, Inc., Upper Saddle River, NJ, USA.

Hong, M.S., Farmayan, W.F., Dortch, I.J., Chiang, C.Y., McMillan, S.K., Schnoor, J.L. (2001). Phytoremediation of MTBE from a groundwater plume. *Environmental Science and Technology*, 35, 1231-1239.

Hsu, F.C., Marxmiller, R.L., Yang, A.Y. (1990). Study of root uptake and xylem translocation of cinmethylin and related compounds in detopped soybean roots using a pressure chamber technique. *Plant Physiology*, 93, 1573-1578.

Kim, J., Drew, M.C., Corapcioglu, M.Y. (2004). Uptake and phytotoxicity of TNT in onion plant. *Journal of Environmental Science and Health, Part A*, 39, 803-819.

Klir, G., Yuan, B. (1995). *Fuzzy sets and fuzzy logic*. Prentice-Hall, Inc. Upper Saddle River, NJ, USA.

Kosko, B. (1997). *Fuzzy engineering*. Prentice-Hall, Inc. Upper Saddle River, NJ, USA.

Krstich, M.A., Schwarz, O.J. (1990). *Characterization of xenobiotic uptake utilizing an isolated root uptake test (IRUT) and a whole plant uptake test (WPUT)*, In: Wang, W., Gorsuch, J., Lower, W. (Eds.), *Plants for Toxicity Assessment*. ASTM International, West Conshohocken, PA.

Limmer, M.A., Burken, J.G. (2014). Plant translocation of organic compounds: molecular and physicochemical predictors. *Environmental Science and Technology*, 1, 156-161.

Liu, Y., Li, Z., Xiong, H., Gao, X., Wu, J. (2010). Understanding of internal clustering validation measures. Data Mining (ICDM), *2010 IEEE International Conference on Data Mining, IEEE*, pp. 911-916.

Manzoni, S., Molini, A., Porporato, A. (2011). Stochastic modelling of phytoremediation. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, https://doi.org/10.1098/rspa.2011.0209.

Miller, E.L., Nason, S.L., Karthikeyan, K., Pedersen, J.A. (2016). Root uptake of pharmaceuticals and personal care product ingredients. *Environmental Science and Technology*, 50: 525-541.

Montgomery, D.C., Peck, E.A., Vining, G.G. (2012). *Introduction to linear regression analysis* (Fifth Ed.). John Wiley & Sons, NJ, USA.

Nazarpour, A., Paydar, G.R., Carranza, E.J.M. (2016). Stepwise regression for recognition of geochemical anomalies: Case study in Takab area, NW Iran. *Journal of Geochemical Exploration*, 168, 150-162.

Orchard, B.J., Doucette, W.J., Chard, J.K., Bugbee, B. (2000). Uptake of trichloroethylene by hybrid poplar trees grown hydroponically in flow-through plant growth chambers. *Environmental Toxicology and Chemistry*, 19, 895-903.

Paraíba, L.C., Queiroz, S.C.N., Maia, A.d.H.N., Ferracini, V.L. (2010). Bioconcentration factor estimates of polycyclic aromatic hydrocarbons in grains of corn plants cultivated in soils treated with sewage sludge. *Science of the Total Environment*, 408, 3270-3276.

Pussemier, L. (1991). Model calculations and measurements of uptake and translocation of carbamates by bean plants. *Chemosphere*, 22, 327-339.

Qiu, J., Chen, G., Xu, J., Luo, E., Liu, Y., Wang, F., Hong Zhou, H., Liu, Y., Zhu, F., Ouyang, G. (2016). In vivo tracing of organochloride and organophosphorus pesticides in different organs of hydroponically grown malabar spinach (Basella alba L.). *Journal of Hazardous Materials*, 316, 52-59.

Qiu, J., Chen, G., Zhou, H., Xu, J., Wang, F., Zhu, F., Ouyang, G. (2016). In vivo tracing of organophosphorus pesticides in cabbage (Brassica parachinensis) and aloe (Barbadensis). *Science of the Total Environment*, 550, 1134-1140.

Ross, T.J. (2009). *Fuzzy logic with engineering applications*. John Wiley & Sons.

Rubin, E., Ramaswami, A. (2001). The potential for phytoremediation of MTBE. *Water Research*, 35, 1348-1353.

Russell, R.S., Shorrocks, V. (1959). The relationship between transpiration and the absorption of inorganic ions by intact plants. *Journal of Experimental Botany*, 10, 301-316.

Samarasinghe, S. (2016). *Neural networks for applied sciences and engineering: from fundamentals to complex pattern recognition*, CRC Press.

San Miguel, A., Ravanel, P., Raveton, M. (2013). A comparative study on the uptake and translocation of organochlorines by Phragmites australis. *Journal of Hazardous Materials*, 244, 60-69.

Sheets, T. (1961). Uptake and distribution of simazine by oat and cotton seedlings. *Weeds*, 9, 1-13.

Shone, M., Bartlett, B., Wood, A.V. (1974). A comparison of the uptake and translocation of some organic herbicides and a systemic fungicide by barley: II. Relationship between uptake by roots and translocation to shoots. *Journal of Experimental Botany*, 25, 401-409.

Sicbaldi, F., Sacchi, G.A., Trevisan, M., Del Re, A.A. (1997). Root uptake and xylem translocation of pesticides from different chemical classes. *Pest Management Science*, 50, 111-119.

Sofizadeh, A., Rassi, Y., Vatandoost, H., Hanafi-Bojd, A.A., Mollalo, A., Rafizadeh, S., Akhavan, A.A. (2016). Predicting the Distribution of Phlebotomus papatasi (Diptera: Psychodidae), the Primary Vector of Zoonotic Cutaneous Leishmaniasis, in Golestan Province of Iran Using Ecological Niche Modeling: Comparison of MaxEnt and GARP Models. *Journal of Medical Entomology*, 54, 312-320.

Su, Y.H., Liu, T., Liang, Y.C. (2010). Transport via xylem of trichloroethylene in wheat, corn, and tomato seedlings. *Journal of Hazardous Materials*, 182, 472-476.

Su, Y.H., Liang, Y.C. (2011). Transport via xylem of atrazine, 2, 4-dinitrotoluene, and 1, 2, 3-trichlorobenzene in tomato and wheat seedlings. *Pesticide Biochemistry and Physiology*, 100, 284-288.

Tanoue, R., Sato, Y., Motoyama, M., Nakagawa, S., Shinohara, R., Nomiyama, K. (2012). Plant uptake of pharmaceutical chemicals detected in recycled organic manure and reclaimed wastewater. *Journal of Agricultural and Food Chemistry*, 60, 10203-10211.

Thompson, P.L., Ramer, L.A., Schnoor, J.L. (1999). Hexahydro-1, 3, 5-trinitro-1, 3, 5-triazine translocation in poplar trees. *Environmental Toxicology and Chemistry*, 18, 279-284.

Trapp, S., Matthies, M., McFarlane, C. (1994). Model for uptake of xenobiotics into plants: validation with bromacil experiments. *Environmental Toxicology and Chemistry*, 13, 413-422.

Trapp, S. (2007). Fruit tree model for uptake of organic compounds from soil and air. *SAR and QSAR Environmental Research*, 18, 367-387.

Undeman, E., Czub, G., McLachlan, M.S. (2009). Addressing temporal variability when modeling bioaccumulation in plants. *Environmental Science and Technology*, 43, 3751-3756.

Von Altrock C. (1995). *Fuzzy logic and neurofuzzy applications explained*. Prentice-Hall, Inc. Upper Saddle River, NJ, USA.

Wold, S., Esbensen, K., Geladi, P. (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 2, 37-52.

Xu, R., Wunsch, D. (2005). Survey of clustering algorithms. *IEEE Transactions on Neural Networks*, 16, 645-678.

Xu, R., Wunsch, D. (2009). *Clustering*. John Wiley & Sons.

Yifru, D.D., Nzengung, V.A. (2006). Uptake of N-nitrosodimethylamine (NDMA) from water by phreatophytes in the absence and presence of perchlorate as a co-contaminant. *Environmental Science and Technology*, 40, 7374-7380.

Yoon, J.M., Oh, B-T., Just, C.L., Schnoor, J.L. (2002). Uptake and leaching of octahydro-1, 3, 5, 7-tetranitro-1, 3, 5, 7-tetrazocine by hybrid poplar trees. *Environmental Science and Technology*, 36, 4649-4655.

Zadeh, L.A., 1965. Fuzzy sets. *Information and Control*, 8, 338-353.

II. EXAMINING PLANT UPTAKE AND TRANSLOCATION OF EMERGING CONTAMINANTS USING MACHINE LEARNING: IMPLICATIONS TO FOOD SECURITY

Majid Bagheri, Khalid Al-jabery, Donald Wunsch, and Joel G. Burken

Civil, Architectural and Environmental Engineering Department, Missouri University of Science and Technology, Rolla, MO 65409

ABSTRACT

When water and solutes enter the plant root through the epidermis, organic contaminants in solution either cross the root membranes and transport through the vascular pathways to the aerial tissues or accumulate in the plant roots. The accumulation of contaminants in plant roots and edible tissues is measured by root concentration factor (RCF) and fruit concentration factor (FCF). In this paper, 1) a neural network (NN) was applied to model RCF based on physicochemical properties of organic compounds, 2) correlation and significance of physicochemical properties were assessed using statistical analysis, 3) fuzzy logic was used to examine the simultaneous impacts of significant compound properties on RCF and FCF, 4) a clustering algorithm (k-means) was used to identify unique groups and discover hidden relationships within contaminants in various parts of the plants. The physicochemical cutoffs achieved by fuzzy logic for the RCF and the FCF were compared versus the cutoffs for compounds that crossed the plant root membranes and found their way into transpiration stream (measured by transpiration stream concentration factor, TSCF). The NN predicted the RCF with improved accuracy compared to mechanistic models. The analysis indicated that log K_{ow}, molecular weight,

and rotatable bonds are the most important properties for predicting the RCF. These significant compound properties are positively correlated with RCF while they are negatively correlated with TSCF. Comparing the relationships between compound properties in various plant tissues showed that compounds detected in the edible parts have physicochemical cutoffs that are more like the compounds crossing the plant root membranes (into xylem tissues) than the compounds accumulating in the plant roots, with clear relationships to food security. The cluster analysis placed the contaminants into three meaningful groups that were in agreement with the results of fuzzy logic.

1. INTRODUCTION

In recent years, there have been growing concerns about the threat that consuming contaminated crops poses to human health. This contamination is a result of the wide use of pesticides in agriculture, irrigation with wastewater treated by conventional treatment systems, and anthropogenic activities (Braune et al., 2005; Gao and Zhu, 2004). Uptake, accumulation and translocation of emerging and fugitive contaminants (EFCs) in plants are of the most alarming issues (García et al., 2019; Mathews et al., 2014). These concerns are based on the growing detection of many EFCs in human breast milk and blood (Bao et al., 2010; Roosens et al., 2010).

The plants absorb the environmental contaminants from the rhizosphere through their roots. Water and solutes enter the root through the epidermis and cross the root membrane to be transported to aerial tissues via xylem/phloem or accumulate in the plant roots (Miller et al., 2016). The accumulation of environmental contaminants in plant roots

is measured by the root concentration factor (RCF) (Trapp, 2004). The RCF correlated mainly with lipophilicity (log K_{ow}) for neutral compounds, but it has an ambiguous relationship with polar and ionizable compounds (Briggs et al., 1982; Miller et al., 2016). Burken and Schnoor (1998) correlated the RCF to the log K_{ow} for organic contaminants using nonlinear regression and the least square minimization method. These single-parameter relationships based on log K_{ow} estimate root-water partitioning for many compounds. However, there are several compounds left as outliers. Considering the other physicochemical properties of compounds for predicting root-water partitioning can improve understanding of the uptake and translocation of compounds in plants.

In recent years, machine learning techniques have gained interests to be applied in various environmental problems due to their ability in predicting chemical and biological processes (Strempel et al., 2013; Zhao et al., 2008). In a recently published study, (Miller et al., 2019) reported that machine learning models are capable in predicting bioconcentration factors in fish and invertebrates. They found that neural networks (NN) and tree-based learners indicate the best predictive performance. In this paper, NN models were used to predict the RCF based on physicochemical properties of compounds. The compound properties include log K_{ow}, molecular weight (MW), hydrogen bond donor (HBD), hydrogen bond acceptor (HBA), rotatable bonds (RB) and polar surface area (PSA). NN models have been successfully used for the prediction purposes in various problems (Bagheri et al., 2017; Torbati, 2016; Yunus, 2017). In a recent study, Bagheri et al. (2019) using NN considerably improved the accuracy of prediction for transpiration stream concentration factor (TSCF), which describes the efficiency of uptake by plants.

Rossi et al. (2019) also reported high performance of NN for the prediction of cerium oxide nanoparticles and cadmium uptake by Brassica napus plants.

The EFCs in the soil are probably transported to aerial tissues, including the edible parts, through plant evapotranspiration (Davis, 1998; Roche et al., 2018). A number of previous studies have reported that uptake of contaminants by plant roots and translocation to shoots is affected by the physicochemical properties of compounds (Briggs et al., 1987; Chen et al., 2015; Gao and Zhu, 2004; Hussein et al., 2007). In almost all of previous studies, the considered compound property is log K_{ow}. Inoue et al. (1998) examined the uptake of two series of amine bases by barley roots and subsequent transport to shoots based on the log K_{ow} of the compounds. They reported that the behavior of these compounds for uptake and translocation in barley plants can be modelled reasonably. Fruit concentration factor (FCF) is normally used to quantify the accumulation of emerging contaminants in edible tissues (Doucette et al., 2018). In this study, by examining physicochemical cutoffs for the FCF and the RCF based on significant properties, and comparing the results with physicochemical cutoffs for the TSCF, the goal was to improve understanding of contaminants with a higher possibility of accumulating in the edible tissues. The physicochemical cutoffs for the TSCF are available in our recently published paper (Bagheri et al., 2019). The results were also compared with the cutoffs for compounds crossing root membranes, compounds in xylem sap with high TSCF (Limmer and Burken, 2014). Statistical analysis was used to find the significant compound properties and also assess the correlation between the properties and root-water partitioning. The simultaneous impacts of physicochemical properties on the accumulation of compounds in plant roots and edible tissues were investigated by fuzzy logic.

A clustering algorithm (k-means) was utilized to reveal hidden relationships and structures among data samples (Xu and Wunsch, 2009). A comprehensive selection of data was compiled to explore the hidden relationships in the uptake and translocation of the environmental contaminants. Clustering was used to group similar contaminants in plant tissues by considering their physicochemical properties, uptake, and translocation. The clustering results were validated by focusing on the physicochemical properties of compounds in different clusters to assess the features of each group in relation to uptake and translocation.

2. MATERIAL AND METHODS

2.1. DATASET

The dataset used for modeling the RCF and examining the simultaneous impacts of compound properties consists of 355 measurements. To assess the physicochemical cutoffs in the edible parts of the plants, 120 measurements of the FCF were complied from the published data. Our dataset also includes 300 measurement of the TSCF to extracts the important features of each group in relation to uptake and translocation (Supporting Information). The dataset consists of various plant species, chemical compounds, and different experimental procedures to make sure model is general and comprehensive (not plant-specific or compound-specific). The measurements for RCF, FCF, and TSCF were not included when there was no evidence of reaching the steady state, roots were damaged, the calculations were not reliable, metabolism of the parent compound in planta was observed, measurements included metabolites, depletion of the dosing solution was higher

than 50%, or other modes of exposure were included in the measurements (Limmer and Burken, 2014). The dataset contains eight variables that describe the physicochemical properties of chemical compounds and plant species. Six are numeric variables (log K_{ow}, MW, HBD, HBA, RB, and PSA) and two are categorical (plant name chemical compounds). The categorical variables were enumerated using the properties encoding method (Lam et al., 2015; Obafemi-Ajayi et al., 2015).

These six selected compound properties are the important predictors for uptake and translocation studies based on the recent studies (Limmer and Burken, 2014). Moreover, considering the physicochemical cutoffs for drugs crossing intestinal membrane, and blood brain barrier, the selected properties in our study make it possible to compare the physicochemical cutoffs in plants with the cutoffs in mammalian systems. The physicochemical properties of chemical compounds were obtained from available chemical structure databases such as US EPA Chemistry Dashboard, and ChemSpider (Bagheri et al., 2019). It should be noted that log K_{ow} was for purely neutral compounds, and lipophilicity for the ionizable compounds was described by log D. Finally, the data was normalized between [0, 1] to accelerate and improve the performance of the applied machine learning algorithms. The six numeric variables were used to build the predictive NN model for the RCF. These properties were used to examine the simultaneous impacts of compound properties, and physicochemical cutoffs at various plant tissues. For the clustering of contaminants in plant tissues, the two categorical variables were also considered in addition to the six numerical variables.

2.2. NEURAL NETWORK MODEL

NN, an intelligent method loosely inspired by neurobiology, was used to predict the RCF based on the physicochemical properties of the compounds. For extensive discussion and information on NN architectures and applications see (Haykin et al., 2009; Hornik, 1991; Huang and Babri, 1998; Leshno et al., 1993; Samarasinghe, 2016). A typical NN consists of at least three (and often many) layers of interconnected computational nodes. A three-layer NN consists of an input layer, a hidden layer, and an output layer. A single-output NN with M computational nodes in the hidden layer is expressed by Equation (1):

$$y(w,x) = \varphi_{out}(\sum_{i=1}^{M}(W_{i,out} \times x_i) + b_{out}) \quad (1)$$

where, φ_{out} out is the activation function of the output layer, $W_{i,out}$ is the weight of the synapsis between the ith neuron in the hidden layer and the output neuron, b_{out} is the bias of the output neuron, and x_i is the output of each neuron in the hidden layer, which can be calculated by Equation (2):

$$x_i = \varphi_h(\sum_{i=1}^{N}(W_{i,n} \times x_{inp}) + b_n) \quad (2)$$

where, φ_h is the activation function of hidden layer, N is the number of input parameters, $W_{i,n}$ is the weight between the ith input parameter and the mth neuron in the hidden layer, b_n is the bias of mth neuron in the hidden layer, and x_{inp} is the ith input parameter.

In this study, different types of NN such as multilayer perceptron (MLP), and radial basis function were used to predict RCF. Our preliminary analysis showed that MLP is the most accurate NN for the prediction of RCF. MLP is one of the most frequently used NN types, which has been successfully used in a number of problems in various fields. MLP is

a capable NN algorithm to use for the regression and mapping. One of the main limitations of the MLP is that this algorithm may stop at local minima instead of global minima. Another limitation of MLP is the need to determine the number of hidden nodes by the users, which can impact the performance of NN when the users do not have enough experience. Setting the number of hidden nodes too low may result in the model underfitting while setting it too high may result in the model overfitting (Ramchoun et al., 2016).

An MLP-NN contains neurons structured in parallel layers, from the inputs to the output. The input nodes receive the input values and pass them on to the first hidden layer nodes. Each node collects the input from all input nodes after multiplying each input value by a random weight, attaches a random bias to this sum, and passes on the results through a non-linear transformation. This forms the input either for the second hidden layer (in an NN with multiple hidden layer) or the output layer that operates identically to the hidden layer. The output of the network is equal to resulting transformed value from the output node (Shanmuganathan, 2016). To improve the accuracy of prediction, MLP-NN utilize different algorithms to train the networks by adjusting weights and biases. Basically the objective of training patterns is to reduce the global error. In this study, the network was trained using different training algorithm such as Levenberg–Marquardt (LM), incremental back propagation, gradient descent backpropagation, gradient descent with adaptive learning rate backpropagation, and batch back propagation.

The RCF prediction model utilizes six physicochemical properties of the compounds including log K_{ow}, MW, HBD, HBA, RB, and PSA as its inputs. The RCF dataset has 355 samples. 70% of these samples were used for training, 15% for validation,

and 15% for testing. The NN prediction performance was measured using the correlation coefficient (R) and the root mean squared error (RMSE). The NN models for predicting the RCF were performed using MATLAB R2014a.

2.3. MODEL TRAINING PROCEDURE

The best network architecture, using a trial-and-error approach, was found to be an MLP-NN with three layers and 45 nodes in the hidden layer. The MLP-NN applied a feed-forward backpropagation network to the input data. After testing different training algorithms, it was observed that the MLP-NN model trained by the well-known Levenberg–Marquardt (LM) learning algorithm gives a higher accuracy. LM is a fast training function but it requires a lot of memory to run as comparing with other training algorithms. The training algorithm attempts to find the optimized weights so that the error of prediction is minimum. LM like other numeric minimization algorithms follows an iterative procedure to train the NN. The weights and biases in each iteration are generated randomly using appropriate functions. Using a trial-and-error approach to select the best transfer functions between layers, the results indicated that with a hyperbolic tangent sigmoid transfer function (*tansig*) between the input and hidden layers and a linear transfer function (*purelin*) between the hidden and output layers make the best prediction. Figure 1 shows the details regarding the architecture and related hyper parameters of the MLP-NN for the best prediction. It should be noted that to train the MLP-NN, the data was divided into three sets using random indices. The largest set including 70% of all data was used to train the MLP-NN. The other two sets each consisting of 15% of all data were used to validate and test the MLP-NN model. The test dataset (15% of all data) is used for cross-

validate the MLP-NN model and test its predictive capabilities since the network does not have access to test dataset during training process.

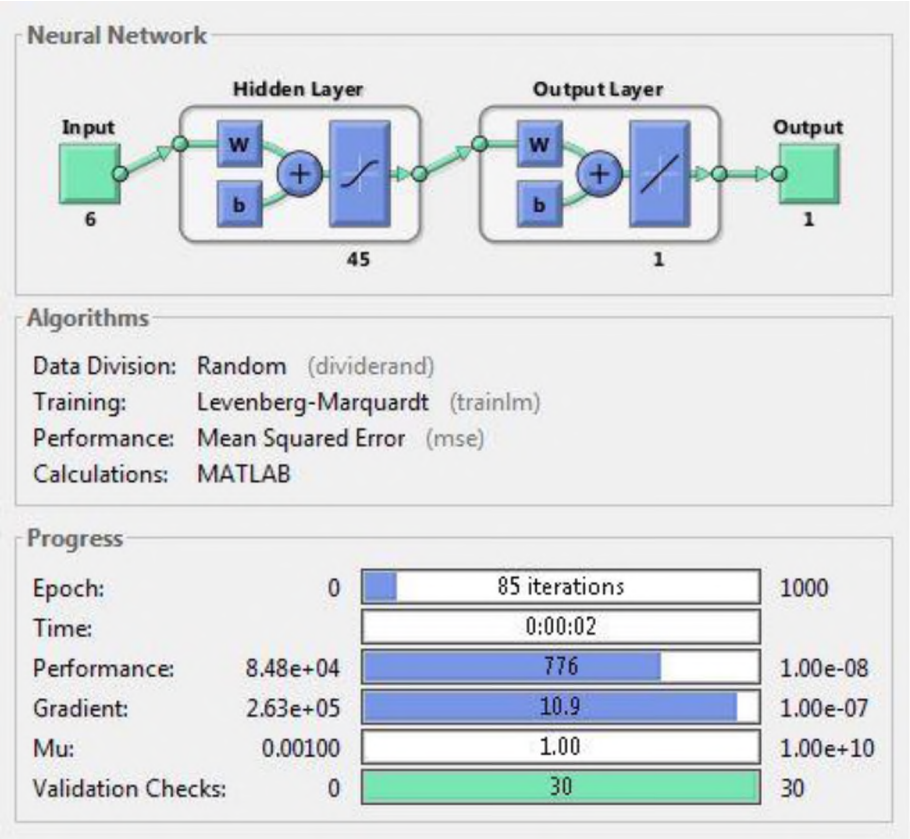

Figure 1. Architecture of the NN and its hyper parameters for the best model.

2.4. STATISTICAL ANALYSIS

Stepwise regression was used to determine the most important predictors in the modeling of root-water partitioning. Stepwise regression is a method by which a predictor may be added or removed from the model at a time to find its importance. It starts with no predictors in the model and calculates the p-value for all predictors not included in the model (Ahlgren and Walberg, 2017). The p-value helps to determine the significance of

the predictors in the model. After adding a predictor (a compound property in this study) to the model, all compound properties in the model are examined to check if their importance has been reduced to a predefined limit of selection (p-critical was 5% in our study). The compound properties with p-values less than 5% are considered significant since the 0.05 was considered as a cutoff for significance. The predictors in the model with importance less than a predefined limit (p-value greater than 0.05) are removed from the model. The chosen predictors were selected based on their p-values and t-test analysis results. The t-statistic is the ratio of the departure of the estimated value of a predictor from its target value to its standard error (Derryberry et al., 2018). This statistical analysis for finding the significant compound properties and their correlation with RCF was performed using IBM SPSS 21.

2.5. FUZZY LOGIC

Fuzzy logic (Zadeh, 1996), was used to examine the partitioning threshold of the chemical compounds through determination of physicochemical cutoffs, and also examining the simultaneous impacts of compound properties on the root-water partitioning. Previous studies have introduced single-parameter relationships for root-water partitioning based on the log K_{ow} of the compounds only (Burken and Schnoor, 1998; Derksen and Keselman, 1992). Fuzzy logic was used in this study because of its ability to handle uncertainty associated with the uptake and the translocation of compounds in plants (Briggs et al., 1982), as well as its ability to investigate the simultaneous impacts of all the chemical compounds. Fuzzy logic consists of four phases: fuzzification, generating fuzzy rules, generating a fuzzy inference system, and defuzzification. Further discussion and

detailed information on fuzzy logic can be found in (Zadeh, 1996). In this intelligent techniques, the real inputs are converted into fuzzified inputs using membership functions. Then fuzzy rules, which are simple *IF-THEN* rules are formed to infer the fuzzy output based on the fuzzified inputs. The *IF-THEN* rules for a system with two inputs of X1 and X2, and one output of Y, can be expressed by Equation (3):

$$IF \quad X_1 \; is \; A \quad and \quad X_2 \; is \; B \quad THEN \quad Y = aX_1 + bX_2 + c \tag{3}$$

where, A and B are membership functions, a, b and c are constants that are adjusted by optimization algorithms.

In the next phase, a fuzzy output is obtained by combining the results of *IF-THEN* rules through a process called inference. Finally, the fuzzy output is converted into a numerical output using membership functions in the defuzzification phase. The analyses using fuzzy logic all were performed in MATLAB R2014a.

2.6. CLUSTERING

The clustering process explores and reveals hidden relationships and structures in uptake and translocation data (Hartigan, 1975; Xu and Wunsch, 2009). Different agglomerative hierarchical clustering and partitional clustering methods were examined to explore the described datasets. The preliminary evaluation showed that the k-means algorithm provides the most compact, isolated and meaningful groups. The k-means algorithm divides data points into a predefined number of clusters k. It initializes the clusters randomly or based on prior knowledge and calculates the clusters' initial matrix $M = [m_1, \ldots, m_k]$. Then, it assigns each data point to the nearest cluster and re-calculates the cluster prototype matrix using the new partitions as in Equation (4).

$$m_i = {}^{1}/_{N_i} \sum_{x_j \in C_i} x_j \quad (4)$$

where, m_i is the sum-of-squared error for cluster i, N_i is the number of elements in cluster i, and x_j is the data point in the cluster C_i.

These steps are repeated until the clusters stop changing. The grouping of the chemical compounds using k-means clustering was performed in MATLAB R2014a. The chemical compounds in different groups for the RCF, the FCF, and the TSCF were compared based their physicochemical properties. The most meaningful grouping was achieved by k=3 because our goal was to compare the physicochemical cutoffs in various plant tissues with the cutoffs obtained by fuzzy logic and published studies.

3. RESULTS AND DISCUSSION

3.1. RCF PREDICTION USING NN

The NN model predicted the RCF with a higher accuracy than those measured with a single parameter and composite models implemented in previous studies (Briggs et al., 1982; Burken and Schnoor, 1998; Topp et al., 1986), as illustrated in Figure 2. According to the training results, the fit between the predicted and measured values for the accumulated contaminants is acceptable, as the regression line for the training dataset demonstrates that the measured and predicted values of the RCF were correlated (Figure 2). The degree of correlation between the predicted values and the measured values was 0.903; together, with an RMSE value of 28.4 for the training set, they confirm the learning capabilities of the NN in the modeling of the RCF. The correlation coefficient and RMSE for the validation model were 0.883 and 29.9, respectively. The degree of correlation

between the predicted values and the measured values for the testing dataset was 0.895, as illustrated in Figure 2. The test error was 31.5, which was even lower than it was for the training and validation sets.

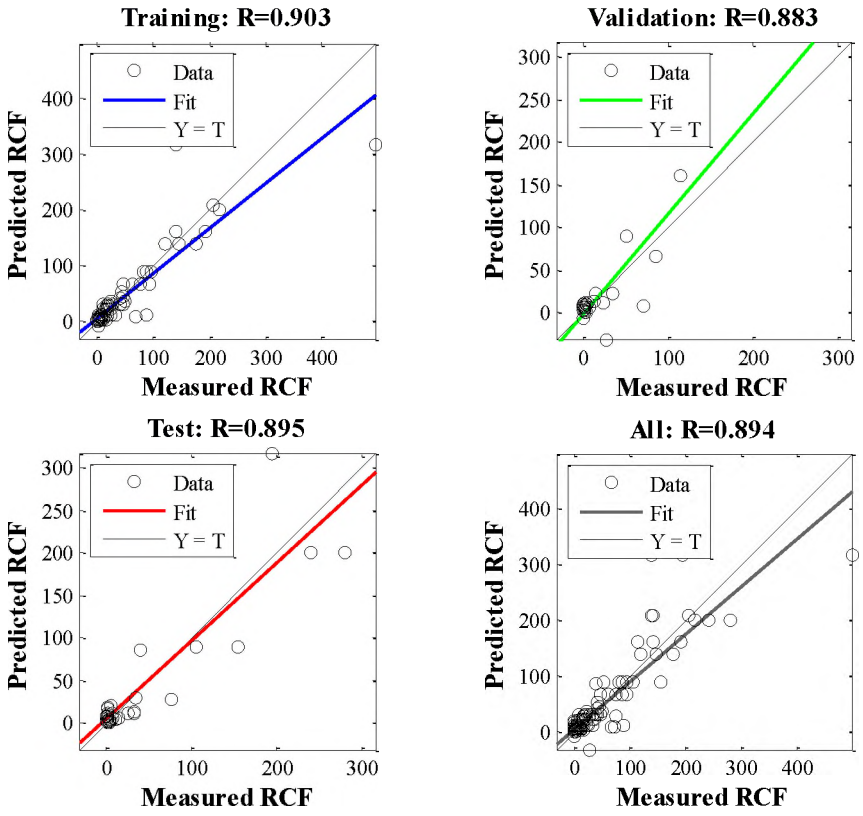

Figure 2. Regression plots of the NN model for root-water partitioning according to training, validation, testing and all datasets.

While in the previous models for predicting RCF do not work well for all compounds and there are some outliers (Briggs et al., 1982; Topp et al., 1986), the NN model predicts root-water partitioning with acceptable error and is applicable for all compounds. The residuals of NN model for train, validation, and test datasets were plotted out against the frequency, see Figure 3. The results indicate a normal distribution of error leading to specific bell-shaped curve, with the highest point in the middle and smoothly

curving symmetrical slopes on both sides of the center. A Gaussian curve (normal distribution of error) reveals our results are symmetrical and approach zero. It was observed that the prediction error is mainly because of moderately hydrophobic compounds with log K_{ow} of 1.42-2.9, and molecular weight of 166-227 Da. The reported RCF values for these compounds vary from 0.5-71. It has been reported in previous studies that the variation in these the uptake and translocation values for these compounds is due to difference in the experimental procedures of uptake studies and experimental artifact (Fantke et al., 2016).

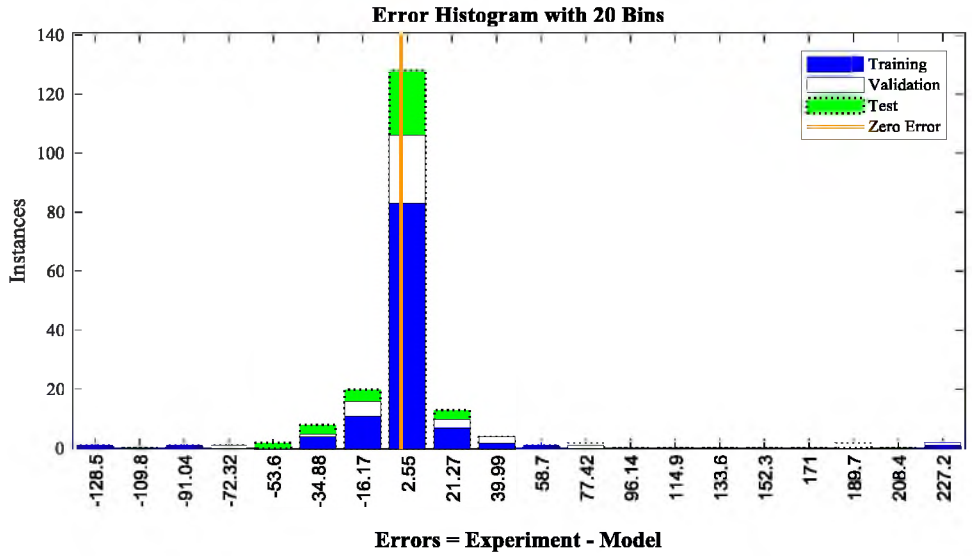

Figure 3. Residual plot (error histogram with 20 bins) for NN model predicting RCF.

NN have shown high accuracy as function approximators in several problems (Bagheri et al., 2019; Rossi et al., 2019; Torbati, 2016). The implemented root-water partitioning model captured the general changes in the RCF values with no oscillations. Traditional partitioning models have limited accuracy because of their single parameter approach, they are unreliable, and their substantial interactions with membrane lipids result

in several outliers (Briggs et al., 1982; Jonker, 2008). The results of this study showed that the NN combined with composite models for root-water partitioning improve understanding of the uptake and translocation of EFCs in plants. The root-water partitioning can also be predicted using poly-parameter linear free energy relationships (pp-LFERs) described by Abraham and others (Abraham et al., 2006; Clarke, 2009). These pp-LFERs have shown high accuracies in predicting partitioning coefficients of environmental models. NN is a capable method, which can accurately predict the partitioning of lipids, cellulose, waxes, hemicellulose, and phenolics to water based on Abraham descriptors. Since the plant root is mainly composed of lipids, cellulose, waxes, hemicellulose, and phenolics, root-water partitioning can be predicted with a high accuracy.

3.2. SIGNIFICANT PROPERTIES AND THEIR CORRELATION WITH RCF

The importance and significance of compound properties including log K_{ow}, MW, HBD, HBA, RB, and PSA for the root-water partitioning was investigated using stepwise regression method. Previous studies mainly considered log K_{ow} as the only significant property, and tried to relate this compound property to the RCF (Briggs et al., 1982). The analysis results from Table 1 show that MW, log K_{ow} and the number of RB are, in order, the most significant compound properties for root-water partitioning. However, the number of RB is not a significant property since the p-value for the RB exceeded 0.05 and the t-statistic was 1.12. Multicollinearity as a common problem was also considered while estimating linear or generalized linear models. Variance inflation factor (VIF) was evaluated as it quantifies how much the variance a coefficient is inflated because of linear

dependence with other predictors. The results indicated that multicollinearity is not a concern in our analysis, and multicollinearity does not exist among significant predictors. For example, the results showed that MW is not collinear with log K_{ow} and RB, as the VIF values are 1.3 and 2.1, respectively. Previous studies have reported HBD as an important properties for the prediction of uptake efficiency. While HBD is not a significant predictor for the root-water partitioning, our analysis indicates that HBD is not collinear with MW and log K_{ow} (VIF less than 3). The lack of collinearity between insignificant properties such as HBA and PSA (VIF of 2.5) conforms that the relative importance of significant predictors are reliable. The correlation analysis also shows that MW, log K_{ow}, and RB are positively correlated with RCF. These results agree with the findings of previous studies, which also consider MW as an important predictor in addition to log K_{ow} (Burken and Schnoor, 1998; Felizeter et al., 2012). The MW was the first significant property with p-value and t-statistic values a little higher than log K_{ow}. It seems that with a larger and more comprehensive dataset, the significance of properties can be investigated better.

Table 1. Pearson correlation of the six compound properties with the RCF.

	RCF	log K_{ow}	MW	HBD	HBA	RB	PSA
RCF	1	0.723	0.779	-0.121	-0.196	0.686	-0.123
log K_{ow}	0.723	1	0.857	-0.315	-0.438	0.656	-0.417
MW	0.779	0.857	1	-0.137	-0.196	0.845	-0.092
HBD	-0.121	-0.315	-0.137	1	0.301	-0.025	0.463
HBA	-0.196	-0.438	-0.196	0.301	1	0.078	0.887
RB	0.686	0.656	0.845	-0.025	0.078	1	0.138
PSA	-0.123	-0.417	-0.092	0.463	0.887	0.138	1

3.3. SIMULTANEOUS IMPACTS AND PHYSICOCHEMICAL CUTOFFS

The results of fuzzy logic indicate that the RCF is proportional with the log K_{ow}, as shown in Figure 4. These findings match the results of previous research (Burken and Schnoor, 1998). The relationship between the log K_{ow} and the RCF was sigmoidal, as illustrated in Figure 4. The RCF was low with the log K_{ow} values of less than 4, and then increased for the log K_{ow} values from 4 to 6. The RCF did not increase considerably for compounds with log K_{ow} values greater than 6. Compounds with higher lipophilicity are more likely to partition to root lipids and concentrate in the plant root because highly lipophilic molecules are not easily transported in the plants even after long term exposure (Miller et al., 2016). The positive correlations between the RCF and root lipid contents has been reported for a number of contaminants in previous studies (Gao and Zhu, 2004). The results show that there is a limited accumulation for hydroophilic compounds in the plant root. More data on the tendency of compounds with higher lipophilicity to accumulate in the plant root can improve our knowledge of uptake and translocation. The RCF also increases significantly with the increase of the MW, based on the compounds' log K_{ow} value. For the compounds with the MW $\geq$ 400 Da, the highest accumulation in the plant root appears when the log $K_{ow} > 4$. For the compounds with MW $\geq$ 400 Da, the increase in accumulation rate is not significant when log $K_{ow} < 4$. This limited increase in accumulation rate is also true for compounds with high lipophilicity (i.e., log $K_{ow} \geq 6$). The accumulation rate in the plant root is more than that for compounds with an MW of less than 400 Da when MW exceeds 400 Da. To assess the accumulation of compounds in plant roots, both the MW and the log K_{ow} should be considered simultaneously. The RB, as the third most important compound property, has a simultaneous effect with the log K_{ow} (Figure 4). The

accumulation in the root increases with the increase of the RB, and the rate of accumulation is more intense when the log $K_{ow} \geq 4$.

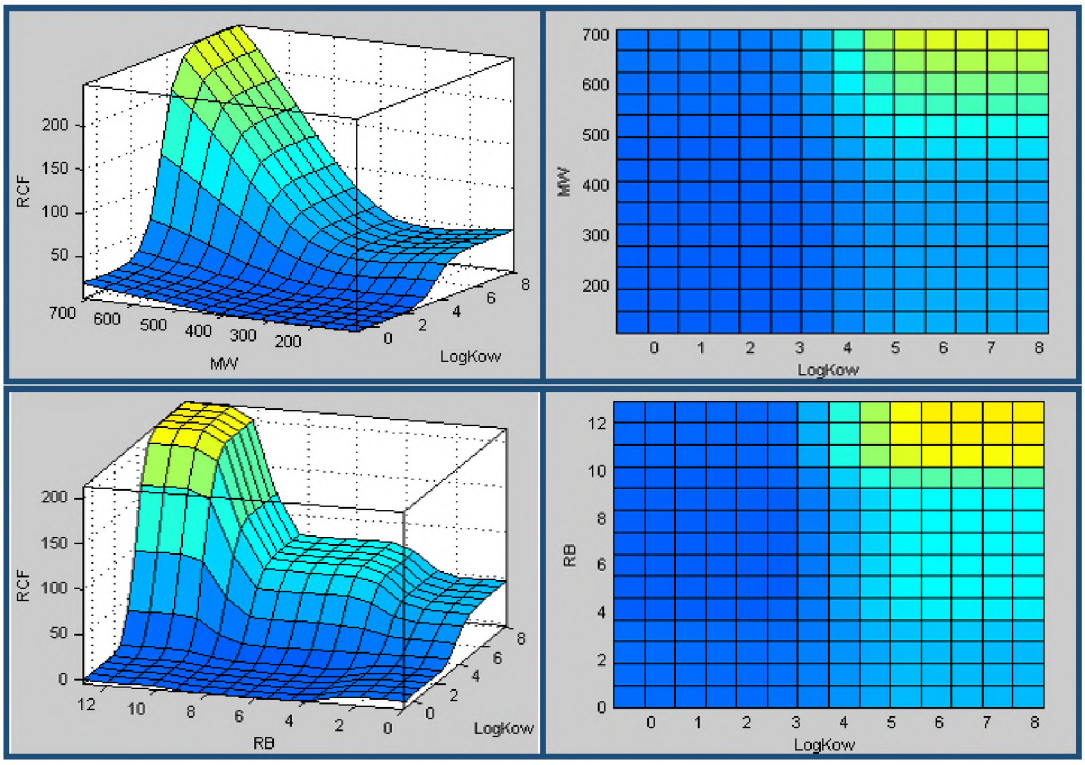

Figure 4. Simultaneous impacts of the log Kow, MW and RB on the RCF, and their physicochemical cutoffs.

This work shows that the MW, log K_{ow}, and RB are positively correlated with the RCF in comparison with the findings of (Bagheri et al., 2019; Limmer and Burken, 2014), on physicochemical cutoffs for compounds with a high TSCF. Limmer and Burken (2014) found that the compounds with log $K_{ow} \leq 4$, MW $\leq$ 350 Da, and RB ≤ 7 are most likely to be translocated by plants. Our results show that the compounds with log $K_{ow} \geq 4$, MW $\geq$ 400 Da, and RB ≥ 8 are most likely to be accumulated in the plant roots (The properties are multiples of four as comparing with the rule of five for the translocatable

compounds). Furthermore, this study indicates that the assessment of the log K_{ow}, MW, and RB determines whether a given compound accumulates in plant roots or is likely to be translocated. To examine a given compound, the physicochemical cutoffs for the three properties should be assessed simultaneously.

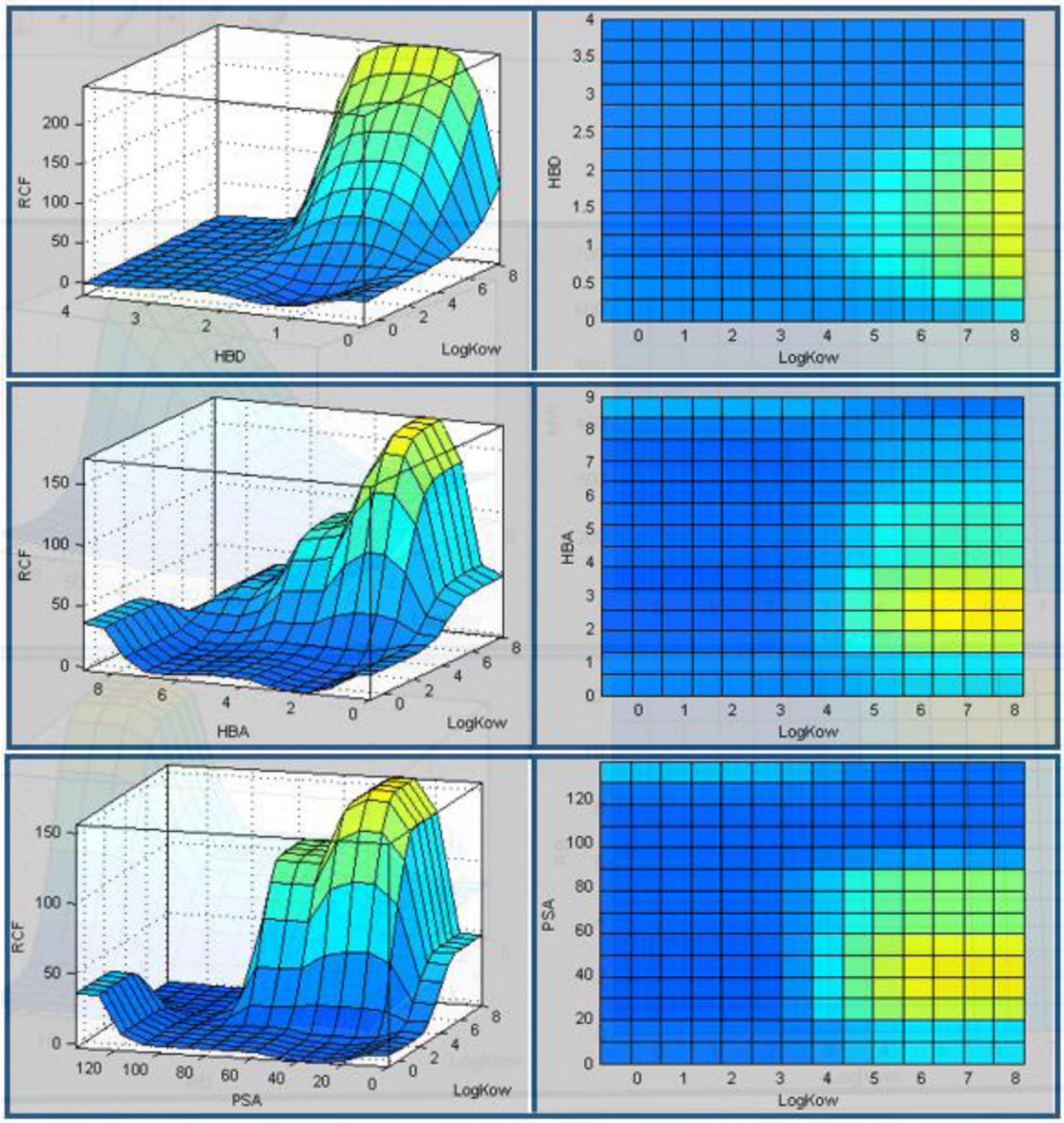

Figure 5. Simultaneous impacts of log Kow and insignificant compound properties (HBD, HBA and PSA) on the RCF, and their physicochemical cutoffs.

The insignificant compound properties including HBD, HBA and PSA are negatively correlated with the RCF, resemble the trend for the translocatable compounds in plants (Limmer and Burken, 2014). As shown in Figure 5, the compounds with HBD <3, HBA $\leq$ 6 and PSA < 80 A2 are more likely to accumulate in the plant root. The simultaneous impacts of these compound properties and log K_{ow} on the accumulation of root-water partitioning are also important here. However, the impacts of these three compound properties are less than the impact of log K_{ow}, MW and RB on the RCF. The physicochemical cutoffs of the HBD, HBA and PSA in this study for the RCF are very similar to those reported for highly translocatable compounds (Bagheri et al., 2019; Limmer and Burken, 2014). The results of this work show that HBD, HBA and PSA are not the appropriate compound properties for examining the likelihood of a given compound to accumulate in a plant root or to be translocated since they share similar cutoffs for both the RCF and the TSCF.

The log K_{ow} and MW are two compound properties with great significance for chemicals that accumulate in plant roots based on the findings of this study, or that cross the plant root membrane based on the findings of (Bagheri et al., 2019; Limmer and Burken, 2014). Therefore, these two compound properties were considered to investigate the uptake and translocation of emerging contaminants in the edible parts of the plants. The results indicated that the relationship between the log K_{ow} and FCF is bell-shaped, and the relationship between the MW and FCF is sigmoidal (Figure 6). By comparing the relationships between the compound properties in various plant tissues, the relationships for the compounds detected in the edible parts is more closely match the compounds crossing plant root membranes (in xylem sap) than compounds accumulating in plant roots.

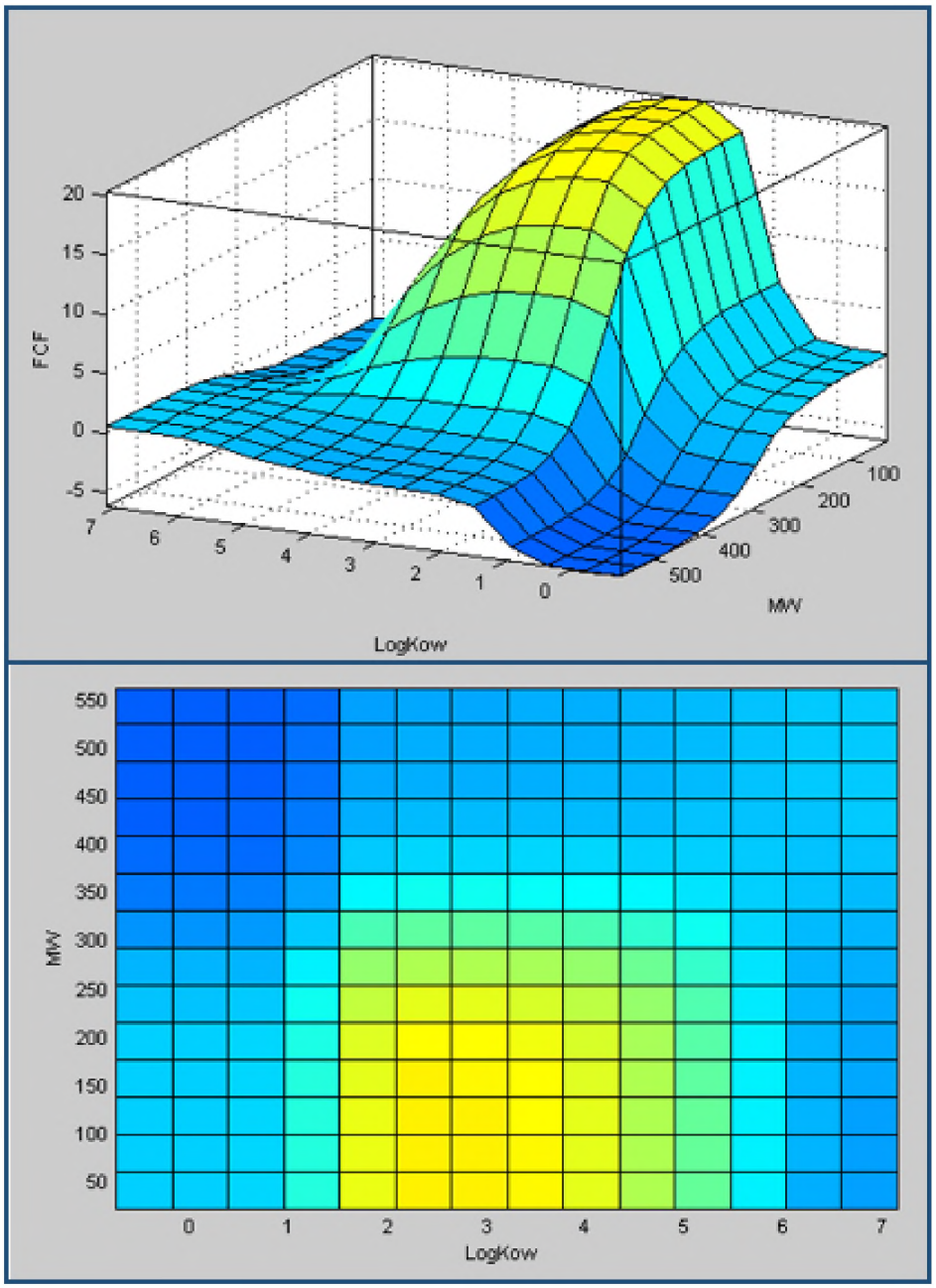

Figure 6. Simultaneous impacts of the log Kow and the MW on the variation of the FCF, and their physicochemical cutoffs.

The compounds with a log K_{ow} < 4 and an MW < 350 Da are more likely to end up in the edible parts of the plants. However, for the compounds with a log K_{ow} < 0 (hydrophilic compounds), the uptake and translocation to the edible parts of the plants

decrease in comparison to moderately hydrophobic compounds (log K_{ow} values of 1-3). For the hydrophilic compounds with an MW $\leq$ 120 Da, the relationship between the log K_{ow} and TSCF is sigmoidal, while the relationship between the log K_{ow} and FCF is bell-shaped. This is interesting as it has been reported in a recent study that hydrophilic compounds are more likely to be taken up and translocated to the edible parts comparing with moderately hydrophobic compounds (Bagheri et al., 2019).

3.4. CLUSTERING OF THE CONTAMINANTS

The k-means algorithm partitioned the compounds into clusters labeled (A), (B) and (C). Cluster (A) contains all the compounds with a high tendency to settle in plant roots. As illustrated in Figure 7, cluster (A) compounds have a log K_{ow} >4 and an MW > 400 Da. These results match those from fuzzy logic, which were discussed earlier. The average RCF for the compounds in this cluster is 143 with log K_{ow} and MW averages of 6 and 580 Da, respectively. The average RCF for cluster (B) compounds is 8.4 (i.e., seventeen times lower than that for (A)) with the log K_{ow} and MW averages of 3.4 and 250 Da, respectively. The average RCF for cluster (C) compounds is 3.76 (i.e., thirty eight times lower than that for (A)) with the log K_{ow} and MW averages of 1.5 and 120 Da, respectively. According to the physicochemical cutoffs introduced by (Limmer and Burken, 2014), using desirability functions, the cluster (B) and (C) compounds are translocatable to aerial tissues. The results indicate that clustering algorithms are capable tools to classify the plant species based on uptake efficiency. One of the important advantages of clustering algorithms is that they can group plant species by considering the physicochemical property of contaminants.

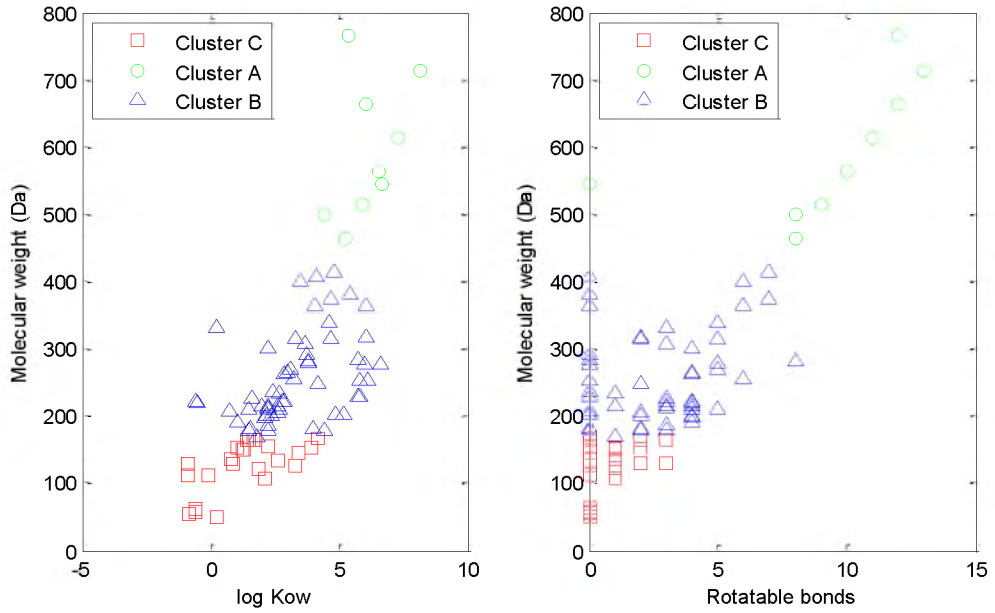

Figure 7. Visualization of the clustered emerging contaminants in plant roots for significant compound properties.

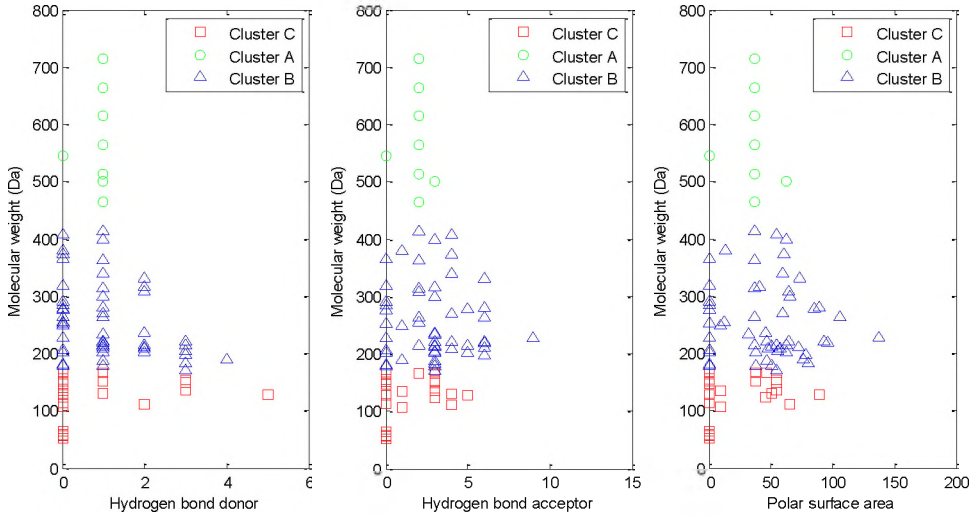

Figure 8. Visualization of the clustered emerging contaminants in plant roots for insignificant compound properties.

The other properties of the compounds in the clusters (i.e., the insignificant properties in terms of accumulation in plant roots) are shown in Figure 8. The results indicate that the MW is the most significant property in compounds classification. The collected results show that the cutoff limits concluded from fuzzy logic are in line with those from k-means. Furthermore, clustering analysis of contaminants in plant roots confirms that the cutoffs for the HBD, HBA and PSA are inversely correlated with RCF.

The most significant clustering criteria for the emerging contaminants in the edible parts of plants was obtained from k-means (k=3). Figure 9 shows the resulting groups and visualizes the compounds in each cluster using MW vs. log K_{ow}. The compounds in each of the three clusters share the same features of physicochemical properties, uptake, and translocation. The average of the physicochemical properties for the compounds in each cluster matches the physicochemical cutoffs achieved by fuzzy logic. Cluster (A) with log K_{ow} and MW averages of 4.8 and 450 Da, respectively has an FCF average of 0.25. The compounds in this cluster share the same physicochemical properties with the compounds that are less likely to cross the plant root membrane. The compounds in cluster (B) have log K_{ow} and MW averages of 3.25 and 250 Da, respectively and the highest FCF average of 11. The compounds in cluster (B) share the same properties with the moderately hydrophobic compounds that have a high likelihood of uptake and translocation, according to the physicochemical cutoffs (Limmer and Burken, 2014). Cluster (C) groups the compounds with a log K_{ow} <0 and an MW<100 Da. The analysis of the compounds in this cluster indicated that the translocation of hydrophilic compounds and heavy metals to the edible parts of the plant is less than that of the moderately hydrophobic compounds. The

compounds in the third cluster have an average log K_{ow}, MW, and FCF of -0.6 and 50 g/mol and 1.2, respectively. These results also match those obtained from fuzzy logic.

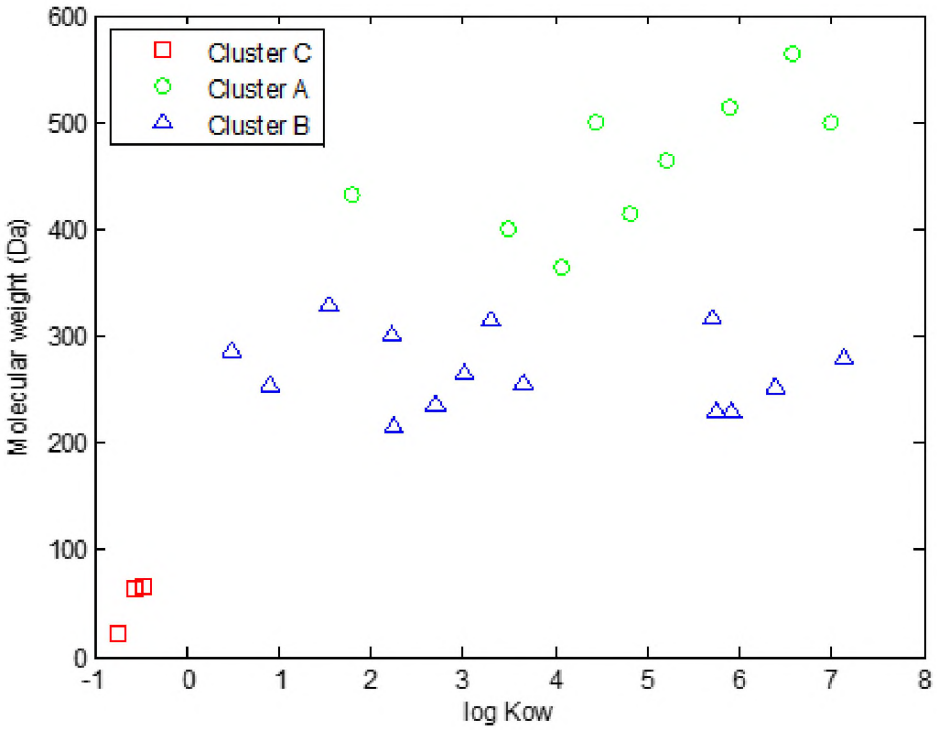

Figure 9. Visualization of the clustered emerging contaminants in the edible parts of the plants.

3.5. HIGH UPTAKE AND IMPLICATIONS TO FOOD SECURITY

Our analysis indicated that perfluorinated compounds are the contaminants with highest FCF. The perfluorinated chemicals with moderate hydrophobicity (log K_{ow} of 2 to 3) have the highest bioconcentration, for some compounds FCF up to 243. This is in a perfect line with the findings of fuzzy logic, the physicochemical properties of the compounds detected in edible parts are similar to those with high TSCF. Moderately hydrophobic compounds have the highest uptake efficiency and TSCF values according to

published studies (Bagheri et al., 2019; Limmer and Burken, 2014). Such high bioconcentration factors for perfluorinated compounds is of concern for both food security and human health, as these compounds have been detected in tomato, strawberry, pea, and cabbage. Based on U.S. environmental protection agency (EPA), perfluorinated chemicals are toxic compounds that includes per- and polyfluoroalkyl substances (PFAS). Some of these compounds are very persistent in the environment and in the human body. They accumulate over time and do not break down; thus, long-term human exposure to them through food consumption is of high concern. Based on U.S. EPA, there is evidence that human exposure to PFAS can lead to adverse health effects. The prediction of bioconcentration factors using our machine learning models is important, especially for untested compounds. U.S. food and drug administration has reported that PFAS contamination in food can occur through environmental contamination, where food are grown in contaminated water and soil. Our analysis also showed that some of the perfluorinated compounds and persistent organic pollutants have the highest RCF values (up to 498). The detected compounds with high RCFs have log K_{ow} of 4 or higher, which is in a good agreement with the physicochemical cutoffs obtained by fuzzy logic. Similar to perfluorinated compounds, persistent organic pollutants are of high concern because they exhibit high lipid solubility, high bioaccumulation in fatty tissues, and high stability. While compounds with higher log K_{ow} have lower capacity for translocation to aerial tissues their high partitioning to root vegetables is a concern for food security and also human health.

4. CONCLUSIONS

Various machine learning techniques were used to: model the accumulation of emerging contaminants in plant roots, compare the physicochemical cutoffs for contaminants in plant roots and edible parts, and examine the capacity of environmental contaminants to reside in various parts of the plants. The results of this study show that: NN is efficient for predicting the environmental contaminants in plant roots (i.e., RCF), and fuzzy logic is efficient in examining the interaction between the physicochemical properties of the environmental contaminants.

The results of fuzzy logic indicated that the physicochemical cutoffs for contaminants in the edible parts of the plants are closer to the cutoffs for compounds crossing plant root membranes (in xylem sap) than the cutoffs for compounds accumulating in the plant roots. Clustering of environmental contaminants using k-means clustering supported the findings of fuzzy logic. The results of k-means produced cutoffs based on molecular weight of chemical compounds that were identical to physicochemical cutoffs obtained from fuzzy logic.

The results of this study show that the integration of machine learning techniques with mechanistic models improves our understanding of the uptake and translocation of the environmental contaminants in plants. The machine learning models can help with improved accuracy of predicting concentrations factors and also other analyses related to the interaction of predictors. Such an integration with mechanistic models can lead to improved understanding of the problem in several aspects.

REFERENCES

Abraham, M.H., Ibrahim, A., Zhao, Y., Acree Jr, W.E. (2006). A data base for partition of volatile organic compounds and drugs from blood/plasma/serum to brain, and an LFER analysis of the data. *Journal of Pharmaceutical Sciences*, 95, 2091-2100.

Ahlgren, A., Walberg, H.J. (2017). *Generalized regression analysis*, in: Bynner, J., Stribley, K.M. (Eds.), *Research Design: The Logic of Social Inquiry*. Routledge, NY, USA.

Bagheri, M., Al-jabery, K., Wunsch, D.C., Burken, J.G. (2019). A deeper look at plant uptake of environmental contaminants using intelligent approaches. *Science of the Total Environment*, 651, 561-569.

Bagheri, M., Bazvand, A., Ehteshami, M. (2017). Application of artificial intelligence for the management of landfill leachate penetration into groundwater, and assessment of its environmental impacts. *Journal of Cleaner Production*, 149, 784-796.

Bao, J., Liu, W., Liu, L., Jin, Y., Dai, J., Ran, X., Zhang, Z., Tsuda, S. (2010). Perfluorinated compounds in the environment and the blood of residents living near fluorochemical plants in Fuxin, China. *Environmental Science and Technology*, 45, 8075-8080.

Braune, B., Outridge, P., Fisk, A., Muir, D., Helm, P., Hobbs, K., Hoekstra, P., Kuzyk, Z., Kwan, M., Letcher, R. (2005). Persistent organic pollutants and mercury in marine biota of the Canadian Arctic: an overview of spatial and temporal trends. *Science of the Total Environment*, 351, 4-56.

Briggs, G.G., Bromilow, R.H., Evans, A.A. (1982). Relationships between lipophilicity and root uptake and translocation of non-ionised chemicals by barley. *Pest Management Science*, 13, 495-504.

Briggs, G.G., Rigitano, R.L., Bromilow, R.H. (1987). Physico-chemical factors affecting uptake by roots and translocation to shoots of weak acids in barley. *Pesticide Science*, 19, 101-112.

Burken, J.G., Schnoor, J.L. (1998). Predictive relationships for uptake of organic contaminants by hybrid poplar trees. *Environmental Science and Technology*, 32, 3379-3385.

Chen, Z.-X., Ni, H.-G., Jing, X., Chang, W.-J., Sun, J.-L., Zeng, H. (2015). Plant uptake, translocation, and return of polycyclic aromatic hydrocarbons via fine root branch orders in a subtropical forest ecosystem. *Chemosphere*, 131, 192-200.

Clarke, E.D. (2009). Beyond physical properties—Application of Abraham descriptors and LFER analysis in agrochemical research. *Bioorganic and Medicinal Chemistry*, 17, 4153-4159.

Corp, I. (2013). *IBM SPSS statistics for windows, version 22.0*. Armonk, NY: IBM Corp.

Davis, L.C.V., S.; Dana, J.; Selk, K.; Smith, K.; Goplen, B.; Erickson, L. E. (1998). Movement of Chlorinated Solvents and Other Volatile Organics through Plants Monitored by Fourier Transform Infrared (FT-IR) Spectrometry. *Journal of Hazardous Substance Research*, 1, 1-26.

Derksen, S., Keselman, H.J. (1992). Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables. *British Journal of Mathematical and Statistical Psychology*, 45, 265-282.

Derryberry, D., Aho, K., Edwards, J., Peterson, T. (2018). Model Selection and Regression t-Statistics. *The American Statistician*, 1-7.

Doucette, W.J., Shunthirasingham, C., Dettenmaier, E.M., Zaleski, R.T., Fantke, P., Arnot, J.A. (2018). A review of measured bioaccumulation data on terrestrial plants for organic chemicals: Metrics, variability, and the need for standardized measurement protocols. *Environmental Toxicology and Chemistry*, 37, 21-33.

Fantke, P., Arnot, J.A., Doucette, W.J. (2016). Improving plant bioaccumulation science through consistent reporting of experimental data. *Journal of Environmental Management*, 181, 374-384.

Felizeter, S., McLachlan, M.S., De Voogt, P. (2012). Uptake of perfluorinated alkyl acids by hydroponically grown lettuce (Lactuca sativa). *Environmental Science and Technology*, 46, 11735-11743.

Gao, Y., Zhu, L. (2004). Plant uptake, accumulation and translocation of phenanthrene and pyrene in soils. *Chemosphere*, 55, 1169-1178.

García, M.G., Fernández-López, C., Polesel, F., Trapp, S. (2019). Predicting the uptake of emerging organic contaminants in vegetables irrigated with treated wastewater–Implications for food safety assessment. *Environmental Research*, 172, 175-181.

Hartigan, J.A. (1975). *Clustering algorithms*, Wiley, NY, USA.

Haykin, S.S. (2009). *Neural networks and learning machines*. Pearson Upper Saddle River, NJ, USA.

Hornik, K. (1991). Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4, 251-257.

Huang, G.-B., Babri, H.A. (1998). Upper bounds on the number of hidden neurons in feedforward networks with arbitrary bounded nonlinear activation functions. *IEEE Transactions on Neural Networks*, 9, 224-229.

Hussein, H.S., Ruiz, O.N., Terry, N., Daniell, H. (2007). Phytoremediation of mercury and organomercurials in chloroplast transgenic plants: enhanced root uptake, translocation to shoots, and volatilization. *Environmental Science and Technology*, 41, 8439-8446.

Inoue, J., Chamberlain, K., Bromilow, R.H. (1998). Physicochemical factors affecting the uptake by roots and translocation to shoots of amine bases in barley. *Pesticide Science*, 54, 8-21.

Jonker, M.T. (2008). Absorption of polycyclic aromatic hydrocarbons to cellulose. *Chemosphere*, 70, 778-782.

Lam, D., Wei, M., Wunsch, D. (2015). Clustering data of mixed categorical and numerical type with unsupervised feature learning. *IEEE Access*, 3, 1605-1613.

Leshno, M., Lin, V.Y., Pinkus, A., Schocken, S., 1993. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. Neural Networks, 6, 861-867.

Limmer, M.A., Burken, J.G. (2014). Plant translocation of organic compounds: molecular and physicochemical predictors. *Environmental Science and Technology*, Letter 1, 156-161.

Mathews, S., Henderson, S., Reinhold, D. (2014). Uptake and accumulation of antimicrobials, triclocarban and triclosan, by food crops in a hydroponic system. *Environmental Science and Pollution Research*, 21, 6025-6033.

Miller, E.L., Nason, S.L., Karthikeyan, K., Pedersen, J.A. (2016). Root uptake of pharmaceuticals and personal care product ingredients. *Environmental Science and Technology*, 50, 525-541.

Miller, T.H., Gallidabino, M.D., MacRae, J.I., Owen, S.F., Bury, N.R., Barron, L.P. (2019). Prediction of bioconcentration factors in fish and invertebrates using machine learning. *Science of the Total Environment*, 648, 80-89.

Obafemi-Ajayi, T., Lam, D., Takahashi, T.N., Kanne, S., Wunsch, D. (2015). Sorting the phenotypic heterogeneity of autism spectrum disorders: A hierarchical clustering model. *IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, pp. 1-7.

Ramchoun, H., Idrissi, M.A.J., Ghanou, Y., Ettaouil, M. (2016). Multilayer Perceptron: Architecture Optimization and Training. *International Journal of Interactive Multimedia and Artificial Intelligence*, 4, 26-30.

R Roche, R.D.L.T., Pagano, L., Majumdar, S., Eitzer, B.D., Zuverza-Mena, N., Ma, C., Servin, A.D., Marmiroli, N., Dhankher, O.P., White, J.C. (2018). Co-exposure of imidacloprid and nanoparticle Ag or CeO2 to Cucurbita pepo (zucchini): Contaminant bioaccumulation and translocation. *NanoImpact*, 11, 136-145.

Roosens, L., D'Hollander, W., Bervoets, L., Reynders, H., Van Campenhout, K., Cornelis, C., Van Den Heuvel, R., Koppen, G., Covaci, A. (2010). Brominated flame retardants and perfluorinated chemicals, two groups of persistent contaminants in Belgian human blood and milk. *Environmental Pollution*, 158, 2546-2552.

Rossi, L., Bagheri, M., Zhang, W., Chen, Z., Burken, J.G., Ma, X. (2019). Using artificial neural network to investigate physiological changes and cerium oxide nanoparticles and cadmium uptake by Brassica napus plants. *Environmental Pollution*, 246, 381-389.

Samarasinghe, S. (2016). *Neural networks for applied sciences and engineering: from fundamentals to complex pattern recognition*. CRC Press, NY, USA.

Shanmuganathan, S. (2016). *Artificial neural network modelling: An introduction*. In: Shanmuganathan S., Samarasinghe S. (Eds) *Artificial Neural Network Modelling (Studies in Computational Intelligence)*. Springer, pp. 1-14.

Strempel, S., Nendza, M., Scheringer, M., Hungerbühler, K. (2013). Using conditional inference trees and random forests to predict the bioaccumulation potential of organic chemicals. *Environmental Toxicology and Chemistry*, 32: 1187-1195.

Topp, E., Scheunert, I., Attar, A., Korte, F. (1986). Factors affecting the uptake of 14C-labeled organic chemicals by plants from soil. *Ecotoxicology and Environmental Safety*, 11, 219-228.

Torbati, S. (2016). Artificial neural network modeling of biotreatment of malachite green by Spirodela polyrhiza: Study of plant physiological responses and the dye biodegradation pathway. *Process Safety and Environmental Protection*, Prot. 99, 11-19.

Trapp, S. (2004). Plant uptake and transport models for neutral and ionic chemicals. *Environmental Science and Pollution Research*, 11, 33-39.

Xu, R., Wunsch, D.C. (2009). *Clustering*. Hoboken. Wiley, NJ, USA.

Yunus, M.A.M. (2017). Artificial Neural Network and Wavelet Features Extraction Applications in Nitrate and Sulphate Water Contamination Estimation. *International Journal of Integrated Engineering*, 9, 64-75.

Zadeh, L.A. (1996). *Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems: Selected Papers by Lotfi A Zadeh*. World Scientific, pp. 394-432.

Zhao. C., Boriani, E., Chana, A., Roncaglioni, A., Benfenati, E. (2008). A new hybrid system of QSAR models for predicting bioconcentration factors (BCF). *Chemosphere*, 73, 1701-1707.

III. INVESTIGATING PLANT UPTAKE OF ORGANIC CONTAMINANTS THROUGH TRANSPIRATION STREAM CONCENTRATION FACTOR AND NEURAL NETWORK MODELS

Majid Bagheri, Xiaolong He, Nadege Oustriere, Wenyan Liu, Honglan Shi, Matt A. Limmer, and Joel G. Burken

Civil, Architectural and Environmental Engineering Department, Missouri University of Science and Technology, Rolla, MO 65409

ABSTRACT

Uptake of seven organic contaminants including bisphenol A, 2,4-dinitrotoluene, estriol, N,N-diethyl-meta-toluamide (DEET), carbamazepine, acetaminophen, and lincomycin by tomato (*Solanum lycopersicum* L.), corn (*Zea mays* L.), and wheat (Triticum aestivum L.) was measured. The plants were grown in a growth chamber under recommended conditions and dosed by these chemicals for 19 days. The plant samples (stem transpiration stream) and solution in the exposure media were taken to measure transpiration stream concentration factor (TSCF). The plant samples were analyzed by a freeze-thaw centrifugation technique followed by high performance liquid chromatography-tandem mass spectrometry detection. Measured average TSCF values were used to test a neural network (NN) model previously developed for predicting plant uptake based on physicochemical properties. The results indicated that moderately hydrophobic compounds including carbamazepine and lincomycin have average TSCF values of 0.43 and 0.79, respectively. The average uptake of DEET, estriol, acetaminophen, and bisphenol A was also measured as 0.34, 0.29, 0.22, and 0.1, respectively. The 2,4-dinitrotoluene was not detected in the stem transpiration stream and it was shown to

degrade in the root zone. Based on these results together with plant physiology measurements, it was concluded that physicochemical properties of the chemicals did predict uptake, however, the role of other factors should be considered in the prediction of TSCF. While NN model could predict TSCF based on physicochemical properties with acceptable accuracies (mean squared error less than 0.25), the results for 2,4-dinitrotoluene and other compounds confirm the needs for considering other parameters related to both chemicals (stability) and plant species (role of lipids, lignin, and cellulose).

1. INTRODUCTION

Scarcity of fresh water is predicted to impact more than 60% of the global population by 2030 (Qadir et al., 2007). Fresh water scarcity in many regions all over the world has resulted in the increased use of water reuse strategies such as reclaiming wastewater for irrigation (Valipour and Singh, 2016). Estimates reveal that over 3.5 million hectares of agricultural land worldwide are irrigated with treated, diluted, partly treated or untreated wastewater (Jimenez and Asano, 2004; Qadir et al., 2010). In addition, biosolids from municipal wastewater facilities and animal manures are frequently used as fertilizers due to multiple economic benefits (Pan and Chu, 2017a), however, biosolids are known to be primary fate of pharmaceuticals and personal care products in both municipal biosolids and animal manures (Wu et al., 2015). Studies over the last two decades confirm that an increasing number of organic compounds are introduced to our agrosystems through the application of biosolids, animal manures, and crop irrigation using reclaimed wastewater (Fu et al., 2019; Wu et al., 2015).

Uptake of antibiotics (tetracycline, sulfamethazine, norfloxacin, erythromycin, and chloramphenicol) by lettuce (*Lactuca sativa*), carrot (*Daucus carota*) and tomato (*Solanum lycopersicum*), and their subsequent translocation to the edible parts indicates that crops are capable of taking up different PPCPs (Pan and Chu, 2017b). Phthalate esters as large-volume chemicals found ubiquitously in soil have been detected in the edible parts of strawberry (*Fragaria* × *ananassa*), lettuce, and carrot (Sun et al., 2015). It was also observed that phthalate esters are metabolized to monoesters in both whole plants and plant cells. Perfluoroalkyl substances and halogenated flame retardants also have been shown to be taken up and translocated by spinach (*Spinacia oleracea*), tomato, and corn (*Zea mays L.*) (Navarro et al., 2017). Human exposure to environmental contaminants through consumption of contaminated crops and vegetables is a pathway of interest in term of environmental and public health. This exposure pathway has received serious attention in recent years due to detection of environmental contaminants in human breast milk and blood (Fromme et al., 2009; Kärrman et al., 2006; Yusa et al., 2012).

The uptake of organic compounds is measured by the transpiration stream concentration factor (TSCF), which is the ratio of compound concentration in the stem transpiration stream to the compound concentration in the exposure media. In this study, the TSCF of seven organic compounds including bisphenol A, estriol, 2,4-dinitrotoluene, N,N-diethyl-meta-toluamide (DEET), carbamazepine, acetaminophen, and lincomycin is investigated. The selected chemicals are a group of compounds with implications for food security, and human health. The United States Food and Drug Administration, European Union and Canada have banned the use of bisphenol A in baby bottles. Several studies have shown that estriol as a medication in menopausal hormone therapy can cause reduced

reproductive capacity and abnormal sexual development with continual long-term exposure at very low concentrations (Karnjanapiboonwong et al., 2011; Kidd et al., 2007). The United States Environmental Protection Agency considers 2,4-dinitrotoluene, which is used to make dyes and explosives as a priority pollutant. Toxicity of this compound to humans and animals has been reported in previous studies (Xu and Jing, 2012). DEET as a compound in insect repellants affects vital cellular functions because it may interact with membrane bound enzymes in mammalian cells (Martinez et al., 2016). In recent studies carbamazepine was detected in vegetables sold commonly in market (Paltiel et al., 2016), and its metabolites were several times higher than the parent compound (Malchi et al., 2014). Acetaminophen is widely used a single-component medication due to its analgesic and antipyretic effects (Ghanem et al., 2016). Chronic acetaminophen exposure to and lincomycin as an antibiotic is important since both pharmaceuticals can alter steroidogenic pathway and increase estrogenicity (Kim et al., 2012).

Accurate prediction of environmental contaminants is important for assessing risks associated with human exposure to these contaminants through food consumption and is a notable gap in current knowledge. Such predictions are fundamental to assess specific compounds as they emerge as contaminants of concern in agro-food systems (Fu et al., 2019). Neural network (NN) modeling is a capable approach to predict the uptake of organic compounds. One aim of this study is to measure the uptake of bisphenol A, estriol, 2,4-dinitrotoluene, DEET, carbamazepine, acetaminophen, and lincomycin by three plant species including tomato, corn and wheat. The chemicals selected in this study cover a wide-range of compound properties, and the three plant species represent a broad botanical space. These compounds are important in terms of human health and the uptake of some

has never been tested in plant systems. The results of uptake experiments are further analyzed to better understand the relationship between TSCF and physicochemical properties of the compounds. The TSCF values for the tested compounds are also predicted by considering their physicochemical properties using NN. Based on these analyses, the implications of these compounds for both food security and human health are discussed.

2. MATERIAL AND METHODS

2.1. CHEMICALS AND REAGENTS

Target standards of estriol (≥97%), 2,4-dinitrotoluene (100 μg/mL in acetonitrile), acetaminophen (≥99%) and bisphenol A (≥99%) were purchased from Sigma-Aldrich (St. Louis, MO, USA). Target standards of carbamazepine (98%), lincomycin (99.9%), and DEET (97%) were purchased from Alfa Aesar (Ward Hill, MA, USA). Table 1 presents the CAS number and physicochemical properties of the compounds used in the uptake experiment. Formic acid (>98%) and ammonium acetate (>99.99%) were purchased from Sigma-Aldrich (St. Louis, MO, USA). Concentrated Hoagland solution was purchased from PhytoTechnology Laboratories (Lenexa, KS, USA). Liquid chromatography–mass spectrometry (LC-MS) grade methanol and acetonitrile were purchased from Fisher Scientific (Fair Lawn, NJ, USA). Ultra-high purity water (18.2 MΩ.cm) was prepared by an Elix-3 water purification system (Millipore, Billerica, MA, USA). Water was degassed and filtered by vacuum filtration using 0.22 μm nylon membrane filter prior to standard preparation and analysis.

Table 1. Physicochemical properties of the compounds used in the uptake experiment.

Compound	CAS Number	log K_{OW}	MW (Da)	HBD	HBA	RB	PSA (A^2)	pKa
Bisphenol A	80-05-7	3.40	228	2	2	4	40	10.2
Estriol	50-27-1	2.90	288	3	3	3	61	10.3
Carbamazepine	298-46-4	2.70	236	2	3	0	46	15.9
2,4-dinitrotoluene	121-14-2	2.10	182	0	6	2	92	13.5
DEET	134-62-3	2.00	191	0	2	3	20	-1.37
Lincomycin	154-21-2	0.91	407	5	8	11	148	7.60
Acetaminophen	103-90-2	0.34	151	2	3	2	49	9.50

Note: Octanol/Water partition coefficient (**log K_{OW}**), molecular weight (**MW**), hydrogen bond donor (**HBD**), hydrogen bond acceptor (**HBA**), rotatable bonds (**RB**), polar surface area (**PSA**), acid dissociation constant (**pKa**).

2.2. PLANT EXPOSURE EXPERIMENT

Tomato (*Solanum lycopersicum* L.) and corn (*Zea mays* L.) seeds were purchased from Johnny's selected seeds (Winslow, ME, USA), and wheat (Triticum aestivum L.) seeds were purchased from Power Grow Systems (Lindon, UT, USA). The seeds were surface-sterilized by sodium hypochlorite solution, and then were rinsed with ultra-high purity (MQ) water before transferring to the Petri dishes with filter papers. A sufficient number of seeds (2-3 times more than the main replicates) were used for germination to provide enough seedlings for each species. Aluminum foil was used to cover the Petri dishes and preserve dark conditions and root temperature. The Petri dishes were checked daily to make sure the filter papers had enough moisture. Germination took 7 days for tomato, and up to 5 days for corn and wheat. The seedlings were then transplanted into the containers (250 ml dark amber glass wide mouth bottles) after the germination step. The containers were filled with glass beads (250-425 microns) to within 2.5 cm from the top of the bottle to support the root apparatus of the seedlings. The same amount of glass beads

was added to all containers to maintain the same conditions for all plants. The plants were grown in a growth chamber with a set temperature of 25 °C, and a set relative humidity of 70% during growing period. The plants were under light intensity of 350 μmol/m^2s, measured at the top of the canopy, using LED lights and a timer to achieve a photoperiod of 16 hours light and 8 hours darkness. The containers on the shelves under the light were positioned randomly, and they were also covered with aluminum foil to avoid evaporation and minimize light in the root zone.

The uptake experiments started with dosing a set of preliminary plants harvesting over time to find the steady state of uptake (exposure time to be used for the main plants). The dosing concentration for these plants harvesting over time was 4 mg/L. This preliminary set consisted of 12 plants, which were harvested at the intervals of 1, 2, 3, 5, 7 (2), 9, 11, 13, 15, 17, and 19 days. The main uptake experiment consisted of three different concentrations with six replicates per treatment level for each group of chemicals (groupings described below). The three concentrations for dosing main plants were 1, 4, and 8 mg/L by considering toxicity of the chemicals and detection limits for the analysis (He et al., 2019). In addition, six replicates were considered as controls (without dosing) to compare the results with the dosed plants for each group, giving a total of 126 containers/plants. The main plants were harvested after finding the steady state conditions. Tomato was the first tested species in the uptake experiments because it probably has higher uptake than wheat and corn (Bagheri et al., 2019). The transpired water was measured by weighing the containers daily for all plant replicates through the exposure period. Plant physiology measurements were recorded to examine the effects of chemicals on the growth of plants.

The control plants were only fed daily with 25% Hoagland solution and no chemicals (He et al., 2019). To feed the controls, pure Hoagland was dissolved in MQ water and a measured volume of the solution was added to each container using an appropriate syringe. To dose the plants with chemicals, the compounds were divided into two groups based on their physicochemical properties, primarily hydrophobicity. The first group of dosing compounds included bisphenol A, estriol, 2,4-dinitrotoluene, carbamazepine, and DEET. The second group included acetaminophen, and lincomycin. Before starting exposure for the main plants, methanol stock solutions were made for all compounds in separate dark mininert vials. The dosing solutions for the three concentrations were made by diluting methanol stock solutions in MQ water. The volume of methanol in the dosing solution was less than 1%. To provide enough nutrients for the plants during the exposure experiments, appropriate amount of pure Hoagland (25% Hoagland) was added to the dosing solution (He et al., 2019). The solution was freshly prepared daily. The dosing solutions (1, 4, and 8 mg/L) were added to the bottom each container using appropriate syringes. New dosing solution was added to each container daily based on the transpired water to keep the solution level and dosing concentrations constant during the exposure.

2.3. SAMPLE PREPARATION

The solution samples from the containers were collected using an appropriate glass pipette immediately before harvesting the plants. The extracted solution from each container was transferred to a tightly sealed vial, and then was kept in refrigerator for analysis. After removing each plant from the containers, the roots were thoroughly rinsed

with deionized water three times. The plant was then divided into three parts including roots, stem, and leaves, however, the stem was analyzed for measuring TSCF. To consider the possible effect of dosing solution on the stem in the containers, 2 cm from the lowest part of stem was omitted and the sample was taken from the upper part. The TSCF was calculated using the transpiration stream solution collected from aerial section of the plant stem. The plant samples (stems) were stored in separate vials and sealed tightly, then were kept in freezer (−20 °C) for subsequent analyses. The internal aqueous solution from the plant stem was separated through a freeze-thaw cycle from plant tissue and macromolecules by a molecular weight cut-off membrane filtration. Briefly, after freezing, the stem samples were thawed at room temperature (20 °C) for up to 60 min to ensure complete thawing. The thawed samples were cut to several pieces and then placed into a membrane centrifugal filter tubes, centrifuged at 13,000 g for 30 min. The liquid filtrate was then transferred into amber glass autosampler vials for analysis (He et al., 2019).

2.4. INSTRUMENTAL ANALYSIS

The analysis of water, and stem transpiration stream from the three plant species was performed by a freeze-thaw centrifugation technique followed by high performance liquid chromatography-tandem mass spectrometry (HPLC-MS/MS) detection (He et al., 2019). A Shimadzu HPLC system (Columbia, MD, USA) has two pumps (LC-20 AD XR), an autosampler (SIL-20AC XR), an online degasser (DGU-30A3), and a column oven (CTO-20A) with a C18 column (75×21 mm, Kinetex. 2.6 µm particle size, Phenomenex, Torrance, CA, USA). Analyst 1.5 software was used for data acquisition and quantification. Carbamazepine, lincomycin, DEET, and acetaminophen were analyzed by positive

electrospray ionization (ESI) mode of HPLC-MS/MS. Estriol, 2,4-dinitrotoluene, and bisphenol-A, were analyzed by negative ESI mode of HPLC-MS/MS. For positive ionization mode, samples were eluted with a flow rate set to 0.25 ml/min under a gradient elution program with eluent A (ultra-pure water with 0.1% (v/v) formic acid) and eluent B (Acetonitrile with 0.1% (v/v) formic acid). For negative ionization mode, samples were eluted with a flow rate set to 0.25 ml/min under a gradient elution program with eluent A (1 mM ammonium acetate in ultra-pure water) and eluent B (1 mM ammonium acetate in methanol). The elution gradient of positive and negative ionization mode are available in our previous method study (He et al., 2019). For positive and negative ionization mode, the MS operated under positive and negative ESI mode, respectively. The ion source conditions for both positive and negative modes were optimized by flow injection analysis (He et al., 2019).

2.5. STATISTICAL ANALYSIS

Statistical analysis was performed using R software version 3.5.1, and MATLAB software version R2017a. Uptake of the chemicals by plant species was assessed using the TSCF, which was calculated by following Equation (1):

$$\text{TSCF} = \frac{\text{Concentration of the compound in stem xylem solution}}{\text{Concentration of the compound in the exposure media}} \quad (1)$$

Pairwise Wilcoxon Tests adjusted with Bonferroni correction were performed to compare TSCF with the concentrations of the chemicals (bisphenol A, 2,4-dinitrotoluene, estriol, carbamazepine, DEET, acetaminophen, and lincomycin) in both containers and plants. These analyses were performed for the three exposure concentrations (1, 4, and 8 mg/L) and the three plant species (corn, tomato, and wheat). Differences were considered

statistically significant at $p < 0.05$. Originally, a one-way analysis of variance was performed but normality and homoscedasticity of residuals were not met for most of the data sets (Shapiro and Levene's test). Min, 1st quantile, median, 3rd quantile, and max values of TSCF, concentration of compounds in containers and in plants (stem transpiration stream) were measured for all the compound, for the three exposure concentrations and for the three plants species. One way ANOVA was performed to examine the role of transpiration and growth of plants (corn, tomato and wheat) depending on exposure concentrations (1, 4 or 8 mg/L) for the three plant species (corn, tomato, and wheat). Differences were considered statistically significant at $p < 0.05$. Normality and homoscedasticity of residuals were met for all tests. When significant differences occurred between treatments, multiple comparisons of mean values were made using post-hoc Tukey HSD tests. Pearson correlation coefficients were assessed between TSCF and evapotranspiration or growth of corn, tomato, and wheat by considering the three exposure concentrations (1, 4, 8 mg/L) to highlight potential links between increases in evapotranspiration or growth and TSCF values. Changes in concentrations of chemicals in the plants (stem) as compared to the concentrations in the containers, the type of plants (corn, tomato, wheat), and their interaction were analyzed for the compounds using an ANCOVA. It should be noted that there is no analyses for 2,4-dinitrotoluene because it was not detected in both plants and- exposure media.

2.6. PREDICTIVE NEURAL NETWORK MODEL

Nonlinear mapping performance of neural network (NN), which is suitable for projection from one set to another is one of its advantages (MehdipourPicha et al., 2019;

Stengel, 2017). Interconnected computational neurons are used to create the network and make the relationships between output and inputs of a given system, for more information, see (Demuth et al., 2014; Haykin, 2007). A typical NN consists of three layers of interconnected neurons (Saadatmand et al., 2020). A single-output NN with M neurons in the hidden layer is expressed by following Equation (2):

$$Y = f(\sum_{n=1}^{N} W_n X_n + b) \quad (2)$$

where, W_n is the weight vector, X_n is the input vector (n = 1, 2… N), b is the bias, f is the transfer function, and Y is the output.

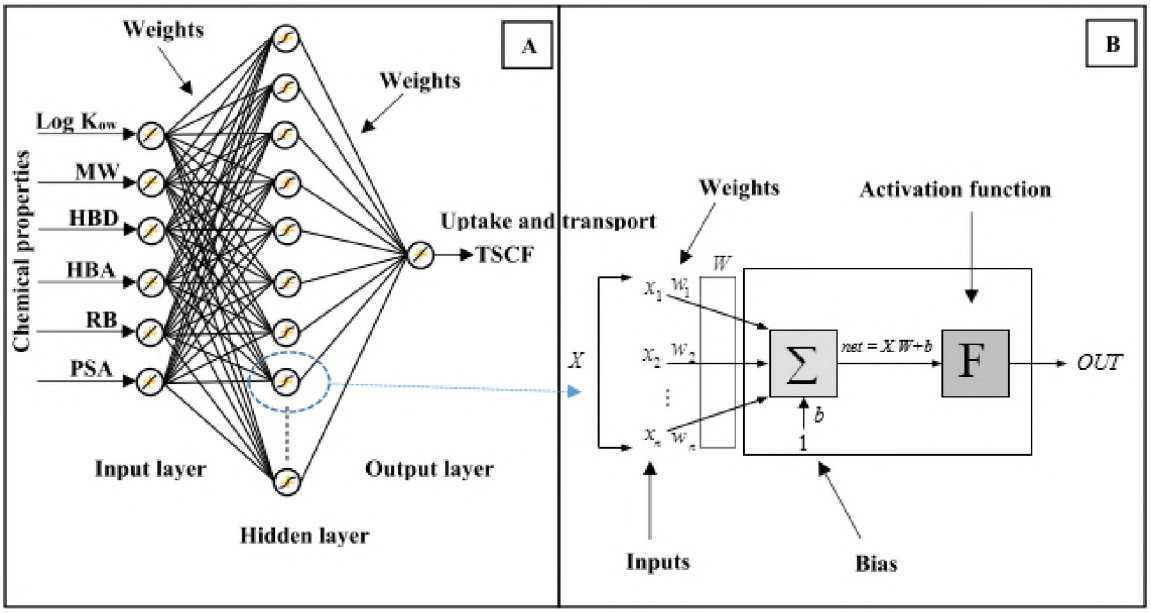

Figure 1. Architecture of the NN predictive model for TSCF (A), architecture of a single computational neuron (B), adapted from (Bagheri et al., 2019).

In this study, multilayer perceptron neural network was used to predict TSCF for the compounds in our experiment. The procedure for NN model development, its architecture, and its training procedure is available in a published paper (Bagheri et al.,

2019). Physicochemical properties of compounds including octanol/water partition coefficient (log K_{ow}), molecular weight (MW), hydrogen bond donor (HBD), hydrogen bond acceptor (HBA), rotatable bonds (RB), and polar surface area (PSA) were used to predict TSCF of the compounds. Figure 1 shows the architecture and inputs of the NN model for the prediction of TSCF.

3. RESULTS AND DISCUSSION

3.1. UPTAKE OF CHEMICALS OVER TIME

The preliminary uptake experiments were performed to find the appropriate exposure time for the main plants. The harvest time of the plants in the primary experiments to measure TSCF is when the uptake of compounds over time in this preliminary experiment reaches a steady state. The changes in the uptake of seven different compounds by the three plant species over time are shown in Figure 2. It was observed that the seven compounds have different trends in uptake over time. For the bisphenol A, the TSCF increases after dosing and then decreases until it reaches to a steady state. This is in agreement with the results of previous studies (Hwang et al., 2017), where the uptake of chlorpyrifos in lettuce grown in soils treated with different concentrations follow a similar trend. The authors claimed that the high concentration in the first days is due to the small weight of lettuce and high residual concentration in soil. Calderón-Preciado et al. (2012) also observed a similar trend for six compounds including four pharmaceuticals and two fragrances. Based on their study, the uptake decreases and reaches to a steady state due to increasing weight of the plant and dissipation of the compound in soil. Our results of the

preliminary experiment indicated that for the estriol, 2,4-dinitrotoluene, acetaminophen, and lincomycin uptake by the plant species reaches its steady state at low TSCF values after few days and then remains constant. This trend for these compounds is potential due to their low uptake, and also from their low stability over time.

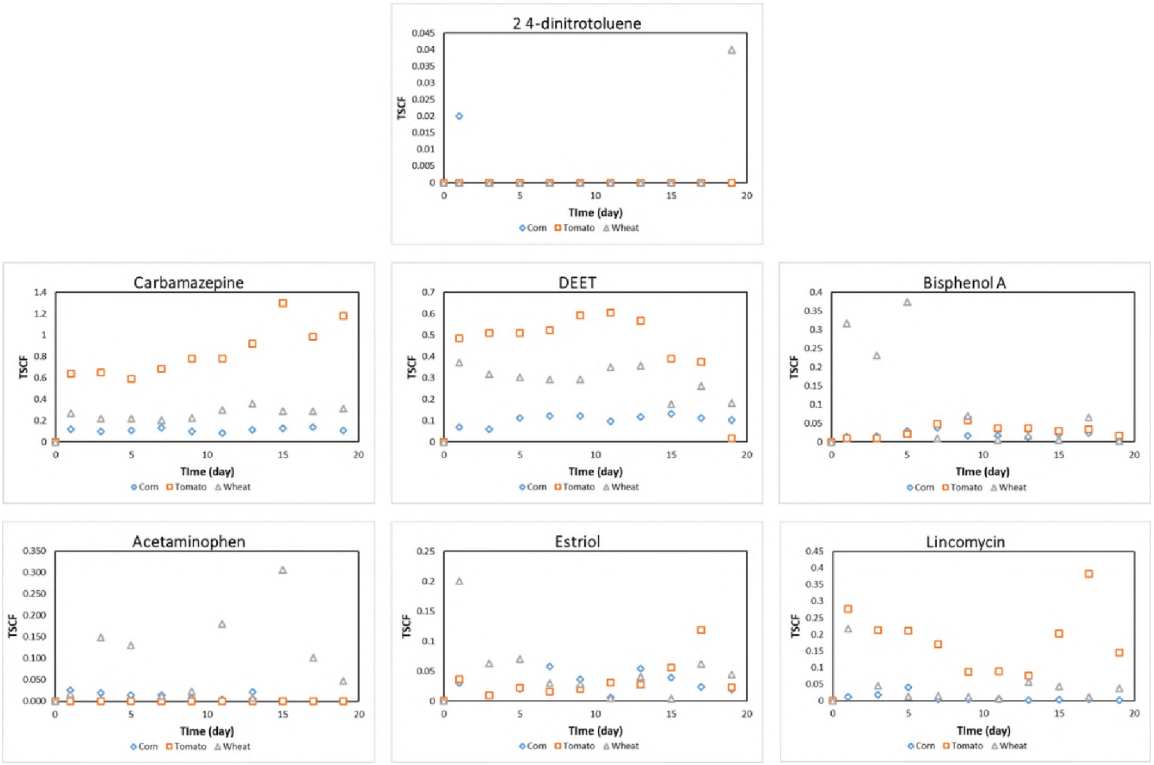

Figure 2. Results of the preliminary experiment to find the steady state of uptake for the seven organic compounds by corn, wheat, and tomato, plants (stem) were harvested at the intervals of 1, 2, 3, 5, 7 (2), 9, 11, 13, 15, 17, and 19 days to measure TSCF (n=1 and dosing concentration of 4 mg/L).

Carbamazepine and DEET demonstrated a different trend as compared with other the compounds. The uptake of these two compounds by corn, wheat and tomato increased constantly for several days, reached to a steady state, and then remained constant. Several organic contaminants such as lindane, hexachlorobenzene, tetrachloroethylene, and

trichloroethylene have been reported following this trend of uptake over time (Li et al., 2005). The results of this preliminary experiment indicated that uptake of the compounds by the three plant species reaches a steady state after 14 days of exposure. While for some compounds the steady state is few days after exposure, the final exposure time was chosen by considering an appropriate exposure time for all chemicals. This is important because plants are dosed with a group of chemicals not a single compound, thus, the final exposure time was chosen by considering the steady state for all compounds.

3.2. VARIATION IN THE UPTAKE BY PLANT SPECIES

The average TSCF values were calculated for the main plants after the preliminary experiment to ensure uptake of the compounds is at a steady state (19 days of exposure). To measure TSCF, the first group of compounds in the main experiments included bisphenol A, estriol, 2,4-dinitrotoluene, carbamazepine, and DEET. Figure 3 shows the TSCF values for these compounds and the three plant species. Similar to preliminary experiments, the concentration of 2,4-dinitrotoluene was lower than detection limit for all of solution samples, and also for almost all of plant samples (stem transpiration stream), for more information about detection limits, see (He et al., 2019). Measuring the concentration of 2,4-dinitrotoluene in various parts of the plants including roots and shoots is necessary to have a better understanding about the fate of this compound. It seems that 2,4-dinitrotoluene is susceptible to degradation prior to possible uptake, as the concentration in both the solution in the root zone solution and plant samples (stem transpiration stream) was not statistically above detection limit, and adsorption to glass beads was not significant. Based on available references, 2,4-dinitrotoluene is only stable

under recommended storage conditions by manufacturers. Su and Liang (2011) reported TSCF values of 0.78 and 1.16 for the uptake of 2,4-dinitrotoluene by wheat and tomato, respectively. However, their experiments were performed under sterilized conditions.

It was observed that uptake (TSCF) of the compounds is different for the three plant species, see Figure 3. The average TSCF for estriol, bisphenol A, carbamazepine, and DEET by tomato was 0.07, 0.02, 0.79, and 0.34, respectively. Among these compounds, carbamazepine is the one that showed highest concentrations in the stem of the three plant species. Carbamazepine is translocated more readily in the plants and accumulates at a higher concentration in transpiring organs (Goldstein et al., 2014). Under dosing concentration of 0.5 μg/L, 24, 29, and 3 ng/g of carbamazepine was detected in cucumber, iceberg lettuce, and spinach leaves, respectively (Wu et al., 2012; Wu et al., 2013). The concentration of estriol and bisphenol A for dosing solution of 1 mg/L in the exposure media decreased at the end of experiment, ranging from 0.019 ± 0.02 (in corn container) to 0.143 ± 0.134 mg/L (in wheat container) for estriol, and from 0.018 ± 0.014 (in tomato container) to 0.078 ± 0.046 mg/L (in wheat container) for bisphenol A. This may originate from the degradation of these compounds in the exposure media, as the concentration of estriol and bisphenol A was also low in the plant samples (stem transpiration stream). The results of a previous study indicate that the concentration of bisphenol A in the exposure media decreases in a greater extent for the plants in non-sterilized conditions than plants in the sterilized conditions (Saiyood et al., 2010). They reported an average uptake of 85% for bisphenol A by Dracaena plant after 20 days, and also showed high partitioning to the roots. Low average TSCF values for both bisphenol A and estriol (0.07 and 0.02) for dosing concentration of 1 mg/L may be due to degradation in the exposure media and also

partitioning to the different parts of the three plant species. Nakajima et al. (2002) found that plant roots are able to absorb bisphenol A and transfer it to leaves. They also found that partitioning of bisphenol A to roots of tobacco is highest compared to stems and leaves. Thus, to reliably measure average uptake of these compounds, the average TSCF values were calculated by considering the dosing concentrations of 4 and 8 mg/L. The average uptake of estriol, bisphenol A, carbamazepine, and DEET by wheat was 0.29, 0.10, 0.36, and 0.07, respectively. The average uptake of estriol, bisphenol A, carbamazepine, and DEET by corn was 0.01, 0.01, 0.18, and 0.14, respectively. Based on the average TSCF values for these compounds, tomato and wheat have a higher uptake than corn.

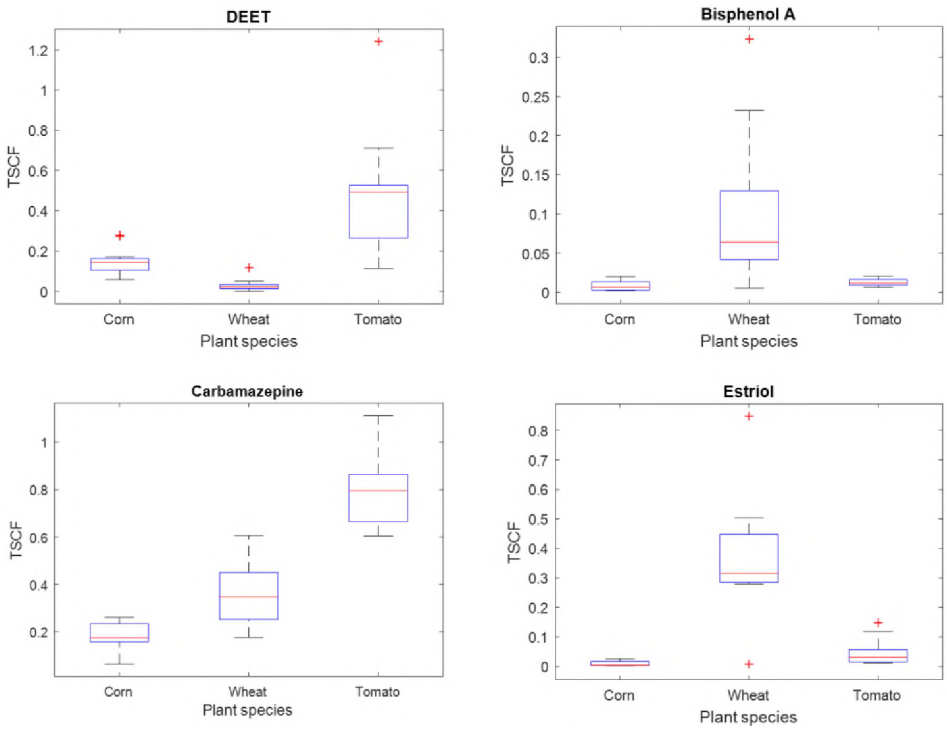

Figure 3. Average uptake (TSCF) of estriol, bisphenol A, carbamazepine, and DEET by corn, wheat and tomato for the main plant samples (n=6 per treatment level) after steady state (19 days of exposure). In the boxplots, the top and bottom of boxes represent the 25th and 75th percentiles, respectively. The line in the middle of each box is the median. The whiskers extend to the most extreme samples not considered outliers. Observations beyond the whisker length are marked as outliers.

Figure 4 shows the uptake (TSCF) of the compounds in the second group including lincomycin and acetaminophen by the corn, wheat, and tomato. The average TSCF for the lincomycin and acetaminophen by corn was 0.01 and 0.21, respectively. For these compounds, the TSCF values for wheat was higher than the TSCF values for corn. The average uptake of lincomycin and acetaminophen by wheat was 0.04 and 0.22, respectively. For lincomycin, concentration of the compound in the exposure media increased significantly due to low uptake and its stability over time. Based on the findings of stability studies for lincomycin (Zhang et al., 2013), this compound maintains its shelf life at 25 °C in aqueous solution with less than 5% degradation occurring over a month. The low TSCF values for lincomycin (0.01 and 0.04) are in agreement with the translocation factor of 0.1 reported for lettuce in a recent study (Chuang et al., 2019). It was also observed that the average uptake of lincomycin by tomato was 0.43. Acetaminophen was not detectable in tomato for both preliminary and main experiments. Since acetaminophen is reasonably a stable compound, the low TSCF values may result from factors such as toxicity to tomato, partitioning to the roots, and accumulation in the upper plant tissues. This can be supported by the results of a recent study, which confirm acetaminophen is taken up by plants (Sun et al., 2019). Bartha et al. (2010) showed that acetaminophen is effectively taken up and translocated to the upper part of Brassica. The results of their study also indicate that after accumulation in the plant roots and shoots, the concentration of acetaminophen decreases significantly. The concentration of acetaminophen in the roots decreased from 1.15 to 0.14 μmol/g, and decreased from 0.71 to 0.07 μmol/g in leaf after seven days.

Considering TSCF values for the seven compounds, it is clear that tomato is the plant species with highest uptake (TSCF values). After tomato, wheat is the second plant species, with higher uptake capabilities as compared with corn. There is no fundamental study on the variability in plant species for the uptake and translocation of contaminants. In a recent study, by applying clustering analysis (a grouping technique) to a data set including various plant species and compounds, based on different thresholds for uptake (different TSCF values from 0.4 to 0.9 to group plants), tomato and wheat mostly showed uptake capacities higher than the various considered thresholds (Bagheri et al., 2019). The findings of this study support our results as they explain that for a given TSCF value as threshold to group plant species, tomato has greater TSCF values than the threshold while corn and wheat show TSCF values less than the threshold.

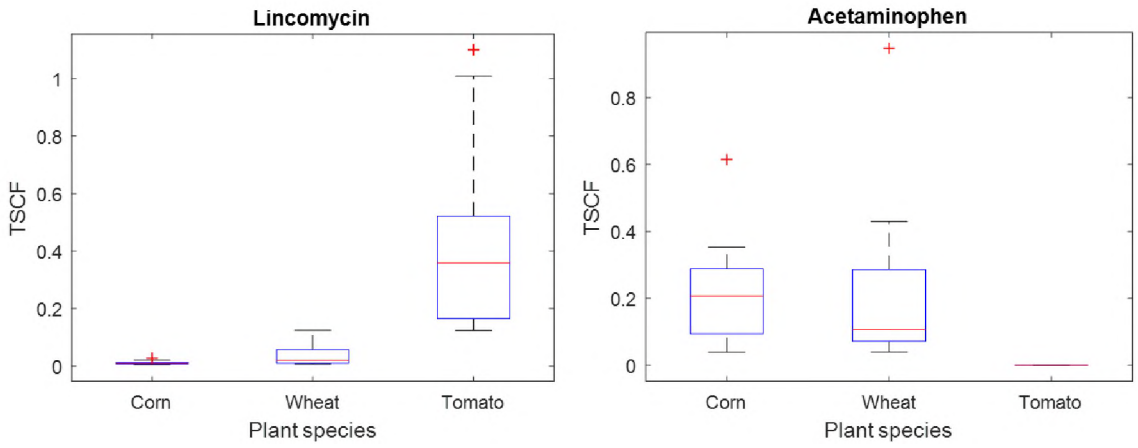

Figure 4. Average uptake (TSCF) of lincomycin and acetaminophen by corn, wheat and tomato for the main plant samples (n=6 per treatment level) after steady state (19 days of exposure). In the boxplots, the top and bottom of boxes represent the 25th and 75th percentiles, respectively. The line in the middle of each box is the median. The whiskers extend to the most extreme samples not considered outliers. Observations beyond the whisker length are marked as outliers.

3.3. EFFECTS OF TREATMENT LEVEL ON PLANT UPTAKE

With increase in the concentration of chemicals in the exposure media, the concentration of the compounds in the plants increases accordingly for the three plant species. Figure 5 shows the concentrations of the first group of the compounds in the exposure media versus their concentrations in the stem of the corn, wheat, and tomato.

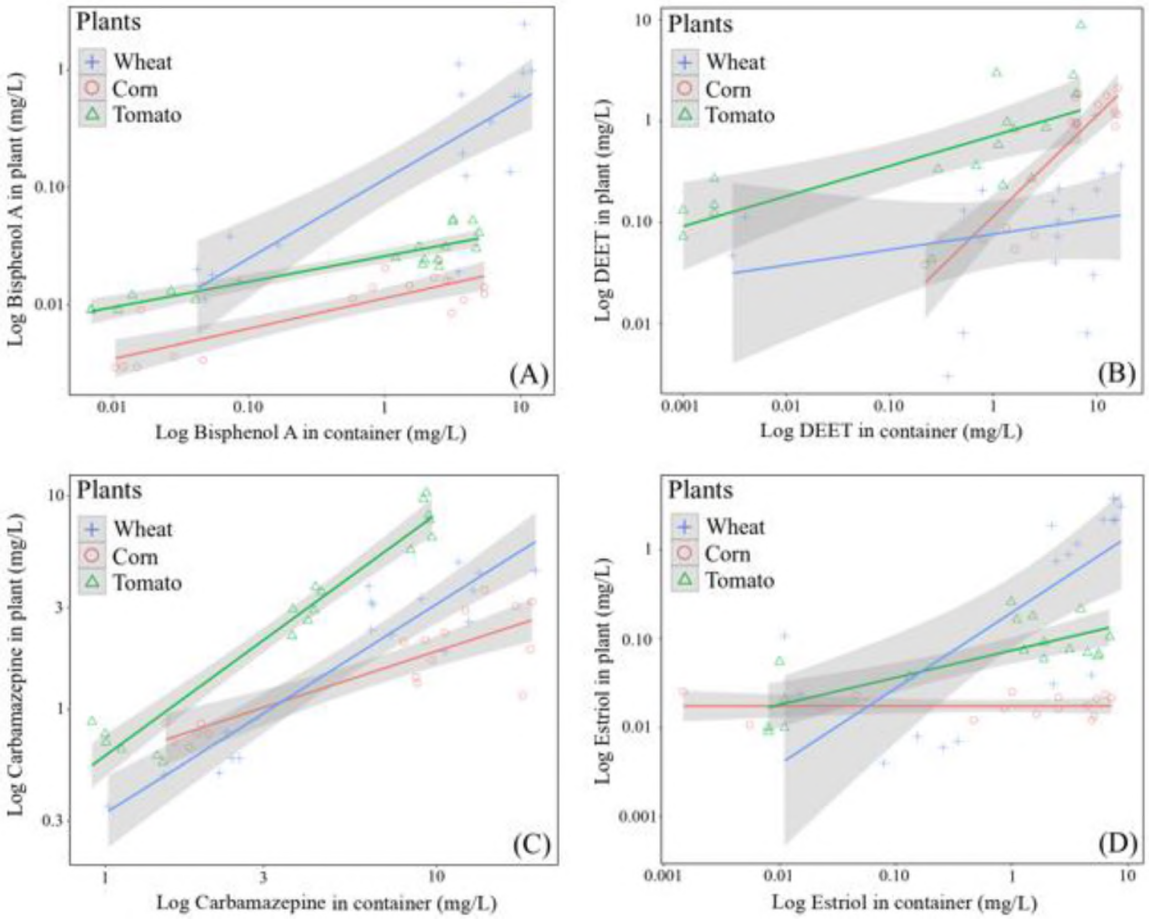

Figure 5. ANCOVA plot for (A) bisphenol A, (B) DEET, (C) carbamazepine, and (D) estriol, in the stem of the plants: wheat (blue), corn (red) and tomato (green) depending on the concentration in the containers (n=6 per treatment level). The trend line power and shaded areas were donated by the ggplot library of the R software. They represent model predictions and corresponding standard error (95%).

While the dosing concentration was the same for all of the compounds and plant species, the detected concentrations in the exposure media were different from the dosing concentration. The difference in the detected concentration in the exposure media for various plant species and chemical compounds probably originates from potential degradation, partitioning to plant roots, and compounds' impact to the plants. This resulted in a wide range of concentrations in the stem of the three plants, ranging from 0.004 to 3.832 mg/L for estriol, 0.003 to 2.459 mg/L for bisphenol A, 0.352 to 10.358 mg/L for carbamazepine, and 0.003 to 8.801 mg/L for DEET, depending on the plants and on the exposure concentration. The measured concentration of estriol, bisphenol A, carbamazepine and DEET in plants depended on the compound concentration in containers, the type of plants, and their interaction. Which means that these two factors (i.e. concentration in containers and type of plants) can have a synergic effect on the concentration of compounds measured in plants (stem transpiration stream), thus, have a synergic effect on the TSCF values (Figure 5 and Table 2). This highlights the needs of understanding more the plants parameters that could influence TSCF.

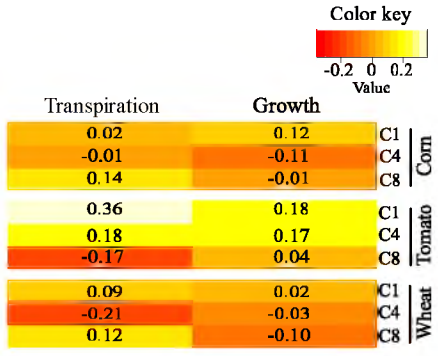

Figure 6. Correlations based on Pearson coefficients between the TSCF and transpiration or growth of corn, tomato and wheat depending on the exposure concentrations (1, 4, and 8 mg/L) of compounds. The C1, C4, and C8 refers to the three dosing concentrations of 1, 4, and 8 mg/L, respectively.

Table 2. Effects of compound concentrations in containers (estriol, bisphenol A, carbamazepine, DEET, acetaminophen and lincomycin), type of plants (corn, tomato, wheat), and their interaction on the concentrations of compounds in plants (stem sap) analyzed by ANCOVA.

Parameters	Df	Mean sq.	F-value	P-value
Estriol				
Conc. in container	1	17.4	106	10^{-12} ***
Plant type	2	5.1	31.2	10^{-8} ***
Conc. in container × Plant type	2	5.9	35.6	10^{-9} ***
Residuals	46	0.2		
Bisphenol A				
Conc. in container	1	4.1	53.5	10^{-8} ***
Plant type	2	0.3	3.4	0.04 *
Conc. in container × Plant type	2	0.3	4.5	0.02 *
Residuals	48	0.1		
Carbamazepine				
Conc. in container	1	53.1	68.6	10^{-10} ***
Plant type	2	46.1	59.5	10^{-12} ***
Conc. in container × Plant type	2	50.8	65.5	10^{-13} ***
Residuals	48	0.8		
DEET				
Conc. in container	1	7.7	9.8	10^{-2} **
Plant type	2	8.3	10.7	10^{-3} ***
Conc. in container × Plant type	2	16.3	20.8	10^{-6} ***
Residuals	48	0.8		
Acetaminophen				
Conc. in container	1	0.36446	84	10^{-9} **
Plant type	1	0.00270	0.6	0.4
Conc. in container × Plant type	1	0.00003	0	0.9
Residuals	32	0.00434		
Lincomycin				
Conc. in container	1	11.4	3.1	0.08
Plant type	2	79.8	21.8	10^{-6} ***
Conc. in container × Plant type	2	14.7	4	0.02 *
Residuals	48	3.7		

Significance levels: '***' 0.001'**' 0.01'*' 0.05; **Df**: degree of freedom; **F-value**: Fisher value; **P-value**: probability value; **Mean sq.**: mean of squares; **Conc. in container**: concentration in container

The results of plant physiology measurements also show that at higher concentrations, 4 and 8 mg/L as compared with 1 mg/L, the transpired water decreases. In this case, for the control plants and the plants dosed with 1 mg/L of the chemicals, the transpired water was about 20 g at the first day and then increased to about 50 g at the last day. In contrast, for the plants dosed with 8 mg/L of the chemicals, the transpired water was about 10 g at the first day, then decreased to 5 g at the final day, see supporting information. The mass of plant replicates at the end of dosing and the changes in the heights also decreased growth of mass for plants dosed with 8 mg/L as compared with the controls. The results indicate that final plant mass and changes in the height is at least 2 times higher for the controls, see supporting information. Previous studies also confirm the toxicity of bisphenol A and estriol to the plants (Adeel et al., 2017; Saiyood et al., 2010). The uptake of bisphenol A and estriol by tomato was less than wheat, probably due to the toxicity of these compounds to tomato, as tomato is the most susceptible species. The transpired water by three plant species, was increasing for corn, decreasing for tomato and wheat over time, see supporting information. Tomato had the sharpest decrease in transpired water probably due to the effects of toxicity at higher concentrations. However, neither the total transpired water nor the growth over time influenced the TSCF for the three plants species and for the three concentrations in the exposure media. The Pearson correlations were performed to assess the effect of transpiration and growth on TSCF. The correlations between -0.21 and 0.36 for all the treatment levels confirm the lack of impact of these two parameters on TSCF, see Figure 6.

The increase in the concentration of lincomycin and acetaminophen in the exposure media, also resulted in an increase of the compounds in plants, for the three plant species,

see Figure 7. For this group of compounds, the original dosing concentration differed from the detected concentration in the exposure media for some of the plant species and the chemical compounds. Such differences can be due to difference in the stability of the compounds, probably partitioning to plant roots, toxicity, and transpired water. For instance, the uptake of lincomycin by tomato was higher than that for wheat and corn. At similar initial exposure concentration, the detected concentration of lincomycin in the exposure media was lower for tomato as compared with wheat and corn, which may indicate the higher transpiration rate of tomato. For both lincomycin and acetaminophen, the measured concentrations in plants (stem transpiration stream) were influenced by concentrations in the containers, and only by the type of plants for the lincomycin. This can be explained by the absence of tomato on the ANCOVA model for the acetaminophen. The difference in the transpiration rate of the three plant species is a factor for the variations of concentration in the exposure media. The compounds in this group, at higher concentrations appeared to trigger some metabolic of toxic response, as evidenced by decrease in the transpired water. The transpiration rate for tomato decreased considerably (up to 70%), the decrease for wheat was also high (up to 40%) while the corn showed a constant transpiration rate over time with no discernable drop. However, the results of plant physiology measurements such as final mass and changes in the height show that the toxic effects of lincomycin and acetaminophen was less than the compounds in the first group, see supporting information. While the degradation of acetaminophen is an important reason for its lower uptake efficiencies, the toxicity is probably another factor for a lack of detection in the tomato. Previous studies confirm the toxicity of acetaminophen to the plant species at higher treatment levels (Nunes et al., 2014).

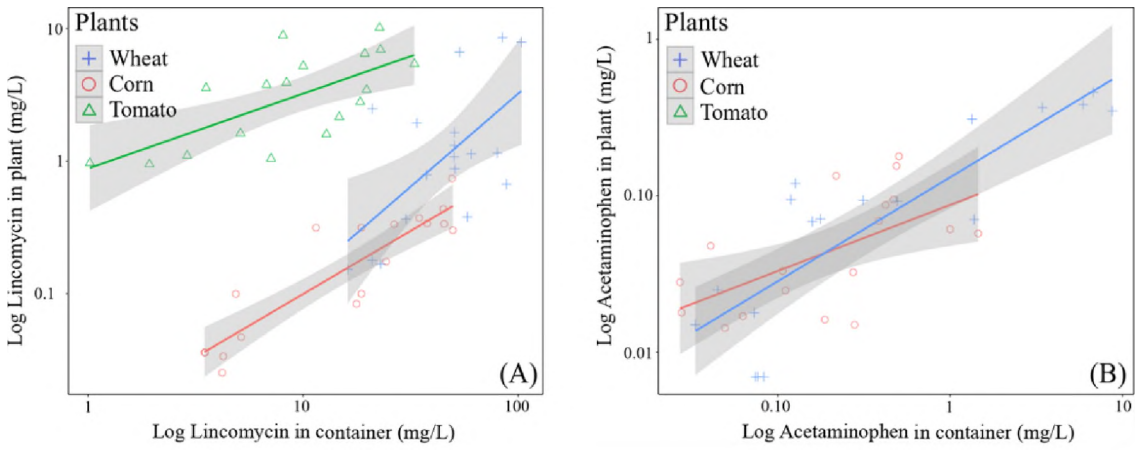

Figure 7. ANCOVA plot for (A) lincomycin and (B) acetaminophen, in the stem of the plants: wheat (blue), corn (red) and tomato (green) depending on the concentration in the containers (n=6 per treatment). The trend line power and shaded areas were donated by the ggplot library of the R software. They represent model predictions and corresponding standard error (95%).

3.4. PLANT UPTAKE AND PHYSICOCHEMICAL PROPERTIES

Numerous previous studies mainly considered log K_{ow} as the significant property to predict plant uptake (De Carvalho et al., 2007; Kim et al., 2004). The findings of these studies indicate that moderately hydrophobic compounds have the highest uptake efficiencies. A recent study also has indicated the high uptake of hydrophilic compounds (Bagheri et al., 2019). Molecular weight, hydrogen bond donor, hydrogen bond acceptor, rotatable bonds, and polar surface area are the other important physicochemical properties (Limmer and Burken, 2014). Carbamazepine as a moderately hydrophobic compound with log K_{ow} of 2.7 showed TSCF values up to 0.79, in close agreement with predicted translocation efficiencies reported for moderately hydrophobic compounds. The uptake of lincomycin as another moderately hydrophobic compound with log K_{ow} of 0.91 is less than carbamazepine. The TSCF values for lincomycin are up to 0.43 as comparing with TSCF

of 0.79 for carbamazepine. Thus, properties other than log K_{ow} are demonstrated to impact the uptake of contaminants. A recent study indicated log K_{ow}, molecular weight, and number of hydrogen bond donors as three significant physicochemical properties with different order of importance (Bagheri et al., 2019). The first significant property (log K_{ow}) was eight times more governing than molecular weight as a predictor. The second significant property (molecular weight) was three times more significant than number of hydrogen bond donors. Low uptake of lincomycin in comparison with carbamazepine can be supported by its higher molecular weight (407 Da vs. 236 Da), and its higher number of hydrogen bond donors (5 vs. 2). This finding is also in agreement with the experimental findings, as the concentration of lincomycin increased in the exposure media due to lower uptake and exclusion at the root membrane.

The results of our experiments indicate that factors other than physicochemical properties play roles in the uptake of organic contaminants. A relatively high transmembrane migration efficiency was expected for 2,4-dinitrotoluene as a moderately hydrophobic compound with log K_{ow} of 2.1, however, its concentration was not only less than detection limits in the plants' stem transpiration stream but also in the exposure media over time following dosing. This further affirms the role of degradation in the uptake of organic contaminants. Previous studies have shown that in-planta concentrations can be used to assess rhizosphere enhanced degradation, and rates of degradation in the bulk soil as well for a range of compounds (Limmer et al., 2018; Wilson et al., 2013). The degradation of acetaminophen, another moderately hydrophobic compound, was also indicated as exposure media concentrations decreases while it was also detected in the plants' transpiration stream. Based on these results, while subsurface degradation probably

is limiting factor for the uptake of acetaminophen, however, translocation in the plant and partitioning to the roots should be studied in future research efforts as acetaminophen has been previously shown to cross root membranes and the average TSCF for this compound was up to 0.22 in our findings as well (Bartha et al., 2010).

In addition, other factors such as toxicity to plants may affect the uptake of these organic contaminants. Tomato was the most susceptible plant species to toxicity based on the plant physiology measurement data. Thus, toxicity and lower transpiration rate might be the reasons for low TSCF values of acetaminophen (0) and estriol (0.07). Similarly, DEET was probably affected by one or combinations of these factors, as the models predict higher TSCF values for this compound (0.61). While bisphenol A should have TSCF values comparable to carbamazepine based on its properties, the possible effects of the above mentioned factors resulted in significantly lower TSCF values for this compound (up to 0.1). This can be supported by the experimental findings, as the concentration of this compound was different in the exposure media for the three plant species, likely due to variable degradation rates.

3.5. TSCF PREDICTIONS BY NN MODEL

The predictions of TSCF using NN model for the compounds in this study were in a general agreement with our understanding about uptake of organic compounds (Bagheri et al., 2019; Limmer and Burken, 2014). The NN could predict the average uptake of bisphenol A, carbamazepine, and estriol with an acceptable accuracy as compared with traditional models (mean squared error less than 0.25), demonstrating one of the first poly-parameter modeling of uptake for emerging pollutants such as pharmaceuticals and

personal care products. The predicted average TSCF for bisphenol A, carbamazepine, and estriol was 0.14, 0.73, and 0.36, respectively. The NN model predicted average TSCF of 0.73 for 2,4-dinitrotoluene, which is in agreement with the results of uptake studies on this compound under sterilized conditions (average TSCF of 0.78-1.16). In our study, due to possible effects of degradation and other environmental conditions, 2,4-dinitrotoluene was not detectable in the plant samples. This shows the importance of other properties such as biodegradability/stability for the uptake of contaminants. To improve the accuracy of predictions, the other parameters related to chemical compounds such as stability over time, and parameters related to plant species such role of fundamental materials (lipids, lignin, and cellulose) should be considered in the modeling. However, these parameters should not be specific to a plant or a chemical because it makes model applicable only for that plant species or chemical. The predicted TSCF for the lincomycin, acetaminophen, and DEET was 0.81, 0.66, and 0.61, respectively as compared with experimental average TSCF values of 0.43, 0.22, and 0.34. While the predicted TSCF values for these compounds are still acceptable (mean squared error less than 0.25), by comparing the predicted values with measured values, it is clear that the role of other factors such as biodegradability/stability should be taken into account. Such considerations are supported by the decreasing concentrations of some of the chemicals in the exposure media over time while for the others, the concentrations increased in the exposure media due to exclusion at the root membrane.

3.6. IMPLICATIONS TO FOOD SECURITY AND HUMAN HEALTH

Uptake and translocation of environmental contaminants can pose risks for food chain security and is associated with human health risks. Among enormous number of environmental contaminants detected in the environment, many have the potential to transport and accumulate in various plant tissues including feedstock and human consumption crops. Bioaccumulation of the environmental contaminants in the edible parts of the plants relates to the physicochemical properties of the contaminants, and other factors such biodegradation in the plants after uptake and translocation, i.e. phytodegradation. Findings support the understanding of physicochemical cutoffs as predictors for the compounds detected in the edible parts of the plants and are similar to the cutoffs for the compounds crossing plant root membranes with high TSCF values (Bagheri et al., 2020). The results of this study show that while moderately hydrophobic compounds have higher tendencies for translocation to the edible parts, the translocation of hydrophilic compounds should also be considered. On the other hand, for the compounds with lower uptake and higher partitioning capabilities to the plant roots (Burken and Schnoor, 1998), the contamination of root vegetables is of higher concern. Carbamazepine as a moderately hydrophobic compound with high uptake is metabolized in several plant species such as tomato, cucumber, sweet potato, and carrot (Riemenschneider et al., 2017). Among the reported metabolites of carbamazepine in plant species, some have a toxic potency even higher than that of the parent compound (Chuang et al., 2018). While degradation of acetaminophen, lincomycin, DEET, and estriol may affect their fate in the environment, and partitioning to the other plant tissues may decrease the potential for bioaccumulation in the edible parts, the risks for food security and human health should

not be neglected. All of these compounds have capacities for translocation to the edible parts according to their physicochemical properties. The findings of the multi-parameter NN derived model give the needed insight for projecting which organic contaminants are most likely to be taken up and should be the targets for uptake studies. Long-term human exposure to these hydrophilic and moderately hydrophobic compounds through food consumption is a potential pathway for human health and understanding the food-relate aspects of our exposome.

4. CONCLUSIONS

In this study, the uptake of seven organic contaminants including bisphenol A, estriol, 2,4-dinitrotoluene, DEET, carbamazepine, acetaminophen, and lincomycin by tomato, wheat and corn was investigated. The analyses indicated carbamazepine and lincomycin as two moderately hydrophobic compounds have average TSCF values of 0.79 and 0.43, respectively. The uptake for the other four compounds including DEET, estriol, acetaminophen and, bisphenol A was 0.34, 0.29, 0.22, and 0.1, respectively. The 2,4-dinitrotoluene was not detected in both plants (stem transpiration stream) and solution samples from exposure media. Considering the wide range of properties for the tested chemicals in this study, it was concluded that physicochemical properties of the compounds can describe the uptake relatively well. However, the effects of degradation, partitioning to various plant tissues, and toxicity should also be considered in future studies. The impact of these factors were confirmed by findings of this study and assessing published literature, as the NN model predicted average TSCF of 0.73, 2,4-dinitrotoluene was not detected in

any of the three plant not the subsurface exposure media possibly due to degradation. The NN model prediction TSCF is close agreement with the TSCF of 0.78 reported for an experiment under sterilized conditions (Su and Liang, 2011). Findings also concluded that tomato was the species with the highest uptake and translocation efficiency, followed by wheat and corn. While this study cannot project a thorough understanding about variability in uptake of chemicals by different plant species or the reasons behind such a variability, the insight gained can lead further research and the enhanced understanding of the multiple physicochemical properties involved will be of great importance in future studies. Addressing these questions is a future research need that will benefit from more precise predictive models.

REFERENCES

Adeel, M., Song, X., Wang, Y., Francis, D., Yang, Y. (2017). Environmental impact of estrogens on human, animal and plant life: a critical review. *Environment International*, 99, 107-119.

Bagheri, M., Al-Jabery, K., Wunsch, D., Burken, J.G. (2020). Examining plant uptake and translocation of emerging contaminants using machine learning: Implications to food security. *Science of the Total Environment*, 698, 133999.

Bagheri, M., Al-jabery, K., Wunsch, D.C., Burken, J.G. (2019). A deeper look at plant uptake of environmental contaminants using intelligent approaches. *Science of the Total Environment*, 651, 561-569.

Bartha, B., Huber, C., Harpaintner, R., Schröder, P. (2010). Effects of acetaminophen in Brassica juncea L. Czern.: investigation of uptake, translocation, detoxification, and the induced defense pathways. *Environmental Science and Pollution Research*, 17, 1553-1562.

Burken, J.G., Schnoor, J.L. (1998). Predictive relationships for uptake of organic contaminants by hybrid poplar trees. *Environmental Science and Technology*, 32, 3379-3385.

Calderón-Preciado, D., Renault, Q., Matamoros, V., Cañameras, N.r., Bayona, J.M. (2012). Uptake of organic emergent contaminants in spath and lettuce: an in vitro experiment. *Journal of Agricultural and Food Chemistry*, 60, 2000-2007.

Chuang, Y.-H., Liu, C.-H., Hammerschmidt, R., Zhang, W., Boyd, S.A., Li, H. (2018). Metabolic demethylation and oxidation of caffeine during uptake by lettuce. *Journal of Agricultural and Food Chemistry*, 66, 7907-7915.

Chuang, Y.-H., Liu, C.-H., Sallach, J.B., Hammerschmidt, R., Zhang, W., Boyd, S.A., Li, H. (2019). Mechanistic study on uptake and transport of pharmaceuticals in lettuce from water. *Environment International*, 131, 104976.

De Carvalho, R.F., Bromilow, R.H., Greenwood, R. (2007). Uptake and translocation of non-ionised pesticides in the emergent aquatic plant parrot feather Myriophyllum aquaticum. *Pest Management Science*, 63, 798-802.

Demuth, H.B., Beale, M.H., De Jess, O., Hagan, M.T. (2014). *Neural network design*. Martin Hagan, Oklahoma, United States.

Fromme, H., Tittlemier, S.A., Völkel, W., Wilhelm, M., Twardella, D. (2009). Perfluorinated compounds–exposure assessment for the general population in Western countries. *International Journal of Hygiene and Environmental Health*, 212, 239-270.

Fu, Q., Malchi, T., Carter, L.J., Li, H., Gan, J., Chefetz, B. (2019). Pharmaceutical and Personal Care Products: From Wastewater Treatment into Agro-Food Systems. *Environmental Science and Technology*, 53, 14083–14090.

Ghanem, C.I., Pérez, M.J., Manautou, J.E., Mottino, A.D. (2016). Acetaminophen from liver to brain: new insights into drug pharmacological action and toxicity. *Pharmacological Research*, 109, 119-131.

Goldstein, M., Shenker, M., Chefetz, B. (2014). Insights into the uptake processes of wastewater-borne pharmaceuticals by vegetables. *Environmental Science and Technology*, 48, 5593-5600.

Haykin, S. (2007). *Neural networks: a comprehensive foundation*. Prentice-Hall, Inc, New Jersey, United States.

He, X., Zhang, H., Xue, R., Liu, W., Bagheri, M., Limmer, M.A., Burken, J.G., Shi, H. (2019). Green Analysis: High Throughput Analysis of Emerging Pollutants in Plant Sap by Freeze–Thaw–Centrifugal Membrane Filtration Sample Preparation—HPLC-MS/MS Analysis. *Journal of Agricultural and Food Chemistry*, 67, 12927-12935.

Hwang, J.-I., Lee, S.-E., Kim, J.-E. (2017). Comparison of theoretical and experimental values for plant uptake of pesticide from soil. *PloS one*, 12, e0172254.

Jimenez, B., Asano, T. (2004). *Acknowledge all approaches: The global outlook on reuse*. Water 21, 32-37, London, UK.

Karnjanapiboonwong, A., Chase, D.A., Canas, J.E., Jackson, W.A., Maul, J.D., Morse, A.N., Anderson, T.A. (2011). Uptake of 17α-ethynylestradiol and triclosan in pinto bean, Phaseolus vulgaris. *Ecotoxicology and Environmental Safety*, 74, 1336-1342.

Kärrman, A., Ericson, I., van Bavel, B., Darnerud, P.O., Aune, M., Glynn, A., Lignell, S., Lindström, G. (2006). Exposure of perfluorinated chemicals through lactation: levels of matched human milk and serum and a temporal trend, 1996–2004, in Sweden. *Environmental Health Perspectives*, 115, 226-230.

Kidd, K.A., Blanchfield, P.J., Mills, K.H., Palace, V.P., Evans, R.E., Lazorchak, J.M., Flick, R.W. (2007). Collapse of a fish population after exposure to a synthetic estrogen. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 8897-8901.

Kim, J., Drew, M.C., Corapcioglu, M.Y. (2004). Uptake and phytotoxicity of TNT in onion plant. *Journal of Environmental Science and Health, Part A*, 39, 803-819.

Kim, P., Park, Y., Ji, K., Seo, J., Lee, S., Choi, K., Kho, Y., Park, J., Choi, K. (2012). Effect of chronic exposure to acetaminophen and lincomycin on Japanese medaka (Oryzias latipes) and freshwater cladocerans Daphnia magna and Moina macrocopa, and potential mechanisms of endocrine disruption. *Chemosphere*, 89, 10-18.

Li, H., Sheng, G., Chiou, C.T., Xu, O. (2005). Relation of organic contaminant equilibrium sorption and kinetic uptake in plants. *Environmental Science and Technology*, 39, 4864-4870.

Limmer, M.A., Burken, J.G. (2014). Plant translocation of organic compounds: molecular and physicochemical predictors. *Environmental Science and Technology*, Letter 1, 156-161.

Limmer, M.A., Wilson, J., Westenberg, D., Lee, A., Siegman, M., Burken, J.G. (2018). Phytoremediation removal rates of benzene, toluene, and chlorobenzene. *International Journal of Phytoremediation*, 20, 666-674.

Malchi, T., Maor, Y., Tadmor, G., Shenker, M., Chefetz, B. (2014). Irrigation of root vegetables with treated wastewater: evaluating uptake of pharmaceuticals and the associated human health risks. *Environmental Science and Technology*, 48, 9325-9333.

Martinez, E., Vélez, S.M., Mayo, M., Sastre, M.P. (2016). Acute toxicity assessment of N, N-diethyl-m-toluamide (DEET) on the oxygen flux of the dinoflagellate Gymnodinium instriatum. *Ecotoxicology*, 25, 248-252.

MehdipourPicha, H., Bo, R., Chen, H., Rana, M.M., Huang, J., Hu, F. (2019). Transformer Fault Diagnosis Using Deep Neural Network. *Innovative Smart Grid Technologies-Asia (ISGT Asia)*. IEEE, pp. 4241-4245.

Nakajima, N., Ohshima, Y., Serizawa, S., Kouda, T., Edmonds, J.S., Shiraishi, F., Aono, M., Kubo, A., Tamaoki, M., Saji, H. (2002). Processing of bisphenol A by plant tissues: glucosylation by cultured BY-2 cells and glucosylation/translocation by plants of Nicotiana tabacum. *Plant and Cell Physiology*, 43, 1036-1042.

Navarro, I., de la Torre, A., Sanz, P., Porcel, M.Á., Pro, J., Carbonell, G., de los Ángeles Martínez, M. (2017). Uptake of perfluoroalkyl substances and halogenated flame retardants by crop plants grown in biosolids-amended soils. *Environmental Research*, 152, 199-206.

Nunes, B., Pinto, G., Martins, L., Gonçalves, F., Antunes, S.C. (2014). Biochemical and standard toxic effects of acetaminophen on the macrophyte species Lemna minor and Lemna gibba. *Environmental Science and Pollution Research*, 21, 10815-10822.

Paltiel, O., Fedorova, G., Tadmor, G., Kleinstern, G., Maor, Y., Chefetz, B. (2016). Human exposure to wastewater-derived pharmaceuticals in fresh produce: a randomized controlled trial focusing on carbamazepine. *Environmental Science and Technology*, 50, 4476-4482.

Pan, M., Chu, L. (2017a). Fate of antibiotics in soil and their uptake by edible crops. *Science of the Total Environment*, 599, 500-512.

Pan, M., Chu, L. (2017b). Transfer of antibiotics from wastewater or animal manure to soil and edible crops. *Environmental Pollution*, 231, 829-836.

Qadir, M., Sharma, B.R., Bruggeman, A., Choukr-Allah, R., Karajeh, F. (2007). Non-conventional water resources and opportunities for water augmentation to achieve food security in water scarce countries. *Agricultural Water Management*, 87, 2-22.

Qadir, M., Wichelns, D., Raschid-Sally, L., McCornick, P.G., Drechsel, P., Bahri, A., Minhas, P. (2010). The challenges of wastewater irrigation in developing countries. gric. *Water Management*, 97, 561-568.

Riemenschneider, C., Seiwert, B., Moeder, M., Schwarz, D., Reemtsma, T. (2017). Extensive transformation of the pharmaceutical carbamazepine following uptake into intact tomato plants. *Environmental Science and Technology*, 51, 6100-6109.

Saadatmand, S., Shamsi, P., Ferdowsi, M. (2020). Power and Frequency Regulation of Synchronverters Using a Model Free Neural Network–Based Predictive Controller. *IEEE Transactions on Industrial Electronics*, DOI: 10.1109/TIE.2020.2984419.

Saiyood, S., Vangnai, A., Thiravetyan, P., Inthorn, D. (2010). Bisphenol A removal by the Dracaena plant and the role of plant-associating bacteria. *Journal of Hazardous Materials*, 178, 777-785.

Stengel, R. (2017). *Introduction to Neural Networks*. Robotics and Intelligent Systems, MAE 345, Princeton University, New Jersey, United States.

Su, Y.H., Liang, Y.C. (2011). Transport via xylem of atrazine, 2, 4-dinitrotoluene, and 1, 2, 3-trichlorobenzene in tomato and wheat seedlings. *Pesticide Biochemistry and Physiology*, 100, 284-288.

Sun, C., Dudley, S., McGinnis, M., Trumble, J., Gan, J. (2019). Acetaminophen detoxification in cucumber plants via induction of glutathione S-transferases. *Science of the Total Environment*, 649, 431-439.

Sun, J., Wu, X., Gan, J. (2015). Uptake and metabolism of phthalate esters by edible plants. *Environmental Science and Technology*, 49, 8471-8478.

Valipour, M., Singh, V.P. (2016). *Global experiences on wastewater irrigation: challenges and prospects*, in: Maheshwari, B., Singh, V. P., Thoradeniya, B. (Eds.), *Balanced urban development: options and strategies for liveable cities*. Springer, Cham, AG, Switzerland, pp. 289-327.

Wilson, J., Bartz, R., Limmer, M., Burken, J. (2013). Plants as bio-indicators of subsurface conditions: impact of groundwater level on BTEX concentrations in trees. *International Journal of Phytoremediation*, 15, 900-910.

Wu, X., Conkle, J.L., Gan, J., 2012. Multi-residue determination of pharmaceutical and personal care products in vegetables. *Journal of Chromatography A*, 1254, 78-86.

Wu, X., Dodgen, L.K., Conkle, J.L., Gan, J. (2015). Plant uptake of pharmaceutical and personal care products from recycled water and biosolids: a review. *Science of the Total Environment*, 536, 655-666.

Wu, X., Ernst, F., Conkle, J.L., Gan, J. (2013). Comparative uptake and translocation of pharmaceutical and personal care products (PPCPs) by common vegetables. *Environment International*, 60, 15-22.

Xu, J., Jing, N. (2012). Effects of 2, 4-dinitrotoluene exposure on enzyme activity, energy reserves and condition factors in common carp (Cyprinus carpio). *Journal of Hazardous Materials*, 203, 299-307.

Yusa, V., Ye, X., Calafat, A.M. (2012). Methods for the determination of biomarkers of exposure to emerging pollutants in human specimens. *TrAC Trends in Analytical Chemistry*, 38, 129-142.

Zhang, D.Q., Hua, T., Gersberg, R.M., Zhu, J., Ng, W.J., Tan, S.K. (2013). Fate of caffeine in mesocosms wetland planted with Scirpus validus. *Chemosphere*, 90, 1568-1572.

SECTION

3. CONCLUSIONS AND RECOMMENDATIONS

3.1. CONCLUSIONS

Environmental contaminants have long been known to cross plant root membranes, and their subsequent translocation to the edible parts is a pathway for potential human exposure to the increasing number of emerging contaminants. Concerns about human health risks associated with the long-term exposure to the emerging contaminants through food consumption are growing and drawing increased attention to food security concerns. Thus, improved understanding of the fate of contaminants in plants is a fundamental step for the risk assessment. Numerous efforts to develop mathematical relationships and relate TSCF to single compound properties (mainly log K_{ow} or Log K_D) in empirical models. However, these relationships generally suffer from low accuracy for many compounds and are limited to specific compounds or plant species. To better understand the fate of emerging contaminants and their relations with human health, new and novel techniques should be employed. Machine learning and statistical analysis are capable tools to investigate the fate of contaminants in the environment, and then to assess the risks associated with human long-term exposure to the contaminants. Using these novel techniques, the importance and role was elucidated for other physicochemical properties such as molecular weight (MW), hydrogen bond donor (HBD), hydrogen bond acceptor (HBA), rotatable bonds (RB), and polar surface area (PSA). These parameters should be considered in assessing potential emerging pollutants that may enter the food chain through

plant uptake. The log K_{ow}, MW, HBD, and RB were the most important compound properties for the uptake of organic contaminants. The log K_{ow} was the most significant property, MW was the second most important, and HBD and RB were two properties with similar importance.

Machine learning is also capable to assess the synergistic impacts of the properties on the uptake and translocation of contaminants in plants. Using fuzzy logic, new relationships were offered between lipophilicity (log K_{ow}) and uptake as described by transpiration steam concentration factor (TSCF). For the compounds with lower molecular weight, the discovered relation between log K_{ow} and TSCF is sigmoidal while for the compounds with higher molecular weight the relation between log K_{ow} and TSCF is bell-shape. The finding clarifies decades of uncertainty and confounding results. This machine learning technique showed high capabilities to investigate translocation of contaminants in plants by obtaining physicochemical cutoffs. It was found that the physicochemical cutoffs for contaminants in the edible parts of the plants are more similar to the cutoffs for compounds crossing plant root membranes (in xylem transpiration stream) than the cutoffs for compounds accumulating in the plant roots. These findings were also supported by clustering of environmental contaminants using k-means clustering algorithm. Through the comparison of the physicochemical cutoffs for the compound crossing plant root membrane and physicochemical cutoffs for drugs crossing intestinal membrane, similarity was uncovered that compounds crossing plant root membrane possibly can also cross intestinal membrane. The physicochemical cutoffs for the plants were actually lower, and therefore more restrictive, than the physicochemical cutoffs for the drugs administered orally.

Modeling and prediction of environmental contaminants in various parts of the plants was also performed using machine learning techniques of neural networks, resulting in improved accuracies. Physicochemical properties of chemical compounds are the most important factors to predict uptake and translocation of emerging contaminants in plants. The most important physicochemical properties relating to distribution were found to be also be lipophilicity, molecular weight, hydrogen bond donors, and rotatable bonds. However, the role of other factors such as plant properties and soil characteristics should be considered to improve the accuracy of predictions. The neural network model predicted the TSCF and RCF with higher accuracy than previous models using physicochemical properties.

To generate new uptake data and also test developed models, various organic chemicals were tested in the uptake experiments. Compounds of unique properties were selected to fulfill unique space in physiochemical properties. Carbamazepine and lincomycin as moderately hydrophobic compounds had measured average TSCFs of 0.79 and 0.43. The TSCFs for DEET, estriol, acetaminophen and, bisphenol A were 0.34, 0.29, 0.22, and 0.1, respectively. The 2,4-dinitrotoluene was not detected in plant stem transpiration stream and solution samples from exposure media. With these chemical compounds and three plant species (tomato, wheat, and corn) a considerable number of TSCF value were generated to test developed models. From the chemicals with unique properties, results indicated that physicochemical properties can describe the uptake relatively well, however, the effects of other factors such as degradation and partitioning to various plant tissues should also be fully considered in future studies. With improved models, variability of plant specific uptake was evaluated. Tomato was the plant species

with the highest uptake for the majority of the unique compounds tested, followed by wheat and corn. While mechanisms of the differences in species was not thoroughly understood, the discovered variability in uptake of chemicals by different plant species was a unique findings and insight gained can be gained from further research.

Partitioning of chemicals between water and plant roots is another factor, which affect the uptake and translocation of environmental contaminants in plants. It is generally believed that chemicals are mainly partitioned to the lipid fraction of plant roots. While lipids are important constituents of plant roots, the role of other fundamental materials such as lignin, cellulose, and protein should not be ignored. The fraction of these fundamental materials in the plant roots is higher than fraction of lipids, which supports considering the role of fundamental materials other than lipids. Developing a model based on all of these fundamental plant materials will increase accuracy of predicting uptake and translocation of environmental contaminants. The results of our uptake experiments also showed that the uptake and translocation is affected by various factors and focusing on one parameter will lead to lower accuracies Degradation of chemicals is a driving factor, which need to be taken into account for the uptake and translocation of contaminants in plants.

To have a better understanding about the possible human health risks associated with long-term exposure to emerging contaminants, the use of new and novel techniques is inevitable. Results also indicate the importance that can be gained by comparing the uptake of chemicals by plants to the uptake of orally administered drugs by human body. The transmembrane migration in these two seemingly dissimilar systems are indeed controlled by the same physiochemical properties. To investigate such similarities, the problem should be examined in both large and molecular scales. Machine learning

techniques such as neural networks, fuzzy logic, and clustering algorithms are some of the greatest tools to investigate similarities in the large scale. With these techniques, potential to compare the physicochemical cutoffs and develop poly parameter models for uptake in both plants and human body. Measuring membrane permeability in both systems is another fundamental issue to gain understanding over the uptake of chemicals by plants and into the human body. Membrane permeability can be measured using simulation techniques examining the interactions of chemicals with biological membranes in a molecular scale.

3.2. RECOMMENDATIONS

The uptake and distribution of emerging contaminants in plants can be correlated with the partitioning of chemicals to the fundamental materials and variation of these materials in different plant tissues. However, understanding is still lacking about the role of fundamental plant materials in the uptake and distribution of emerging contaminants. With the improved understanding of how physicochemical properties are interrelated to control uptake, the impact on partitioning to specific plant materials is within reach. While experimental studies are good, for an improved understanding of the uptake and translocation of chemicals in plants, the use of new and novel approaches is inevitable. Considering the physicochemical cutoffs for organic compounds, there should be similarities between transmembrane permeability in plants and mammalian. New and novel simulation techniques are also capable to examine the similarities between various biological membranes in plant roots and mammalian. To advance the fundamental knowledge of plant uptake and translocation of emerging contaminants, by considering the

importance of physicochemical properties and fundamental plant materials, the following research needs are proposed for the future efforts:

The importance of fundamental plant materials for the distribution of emerging contaminants in plants should be investigated: Lipid fraction of plant roots plays a pivotal role in the uptake of contaminants and their subsequent translocation to the areal parts. The role of other fundamental materials such as lignin, cellulose, and proteins on the uptake of contaminants by plant roots, and their translocation to other parts of the plants worth investigation. The effect of fractions of these fundamental materials in different parts of various plant species and their relation to bioaccumulation of contaminants at various parts of the plants is an open issue for the future studies.

Partitioning of emerging contaminants to fundamental plant materials need to be assessed from a molecular standpoint: Uptake, translocation, and fate of the contaminants in plants are directly related to the partitioning of compounds between water and plant's fundamental materials such as lipid, lignin, and cellulose. Molecular dynamics simulation is a capable approach to provide insight into the fate of emerging contaminants in plants by assessing absorption of contaminants to the fundamental materials in a molecular scale. Future studies should assess variability in plants in terms of uptake and translocation of various emerging contaminants: Variability in the uptake and translocation of emerging contaminants in different plant species has implication for both phytoremediation and food security. Artificial intelligence and machine learning are useful to determine the variability in different plant species. With such assessments one can find the plant species with potential for phytoremediation, and also plant species with higher possibility of contamination in their edible parts.

Comparison between plant uptake of chemical compounds and human uptake of drug-like compounds is also interesting: Knowledge of membrane permeability in mammalian (blood-brain barrier, and membranes in the gastrointestinal tract) and its similarity with plant root membrane is of paramount importance for assessing human health risks, and also in drug development. Artificial intelligence and machine learning are robust techniques to assess the roles of physicochemical properties in the uptake efficiency of chemicals/drugs in plants/humans.

BIBLIOGRAPHY

Briggs GG, Bromilow RH, Evans AA. Relationships between lipophilicity and root uptake and translocation of non-ionised chemicals by barley. Pest Management Science 1982, 13, 495-504.

Burken JG, Schnoor JL. Predictive relationships for uptake of organic contaminants by hybrid poplar trees. Environmental Science & Technology 1998, 32, 3379-3385.

Calderón-Preciado D, Matamoros V, Bayona JM. Occurrence and potential crop uptake of emerging contaminants and related compounds in an agricultural irrigation network. Science of the total environment 2011, 412, 14-19.

Chen L, Zhang S, Huang H, Wen B, Christie P. Partitioning of Phenanthrene by Root Cell Walls and Cell Wall Fractions of Wheat (Triticum aestivum L.). Environmental Science & Technology 2009, 43, 9136-9141.

Chiou CT, Sheng G, Manes M. A partition-limited model for the plant uptake of organic contaminants from soil and water. Environmental science & technology 2001, 35, 1437-1444.

Collins CD, Finnegan E. Modeling the plant uptake of organic chemicals, including the soil− air− plant pathway. Environmental science & technology 2010, 44, 998-1003.

De Carvalho RF, Bromilow RH, Greenwood R. Uptake and translocation of non-ionised pesticides in the emergent aquatic plant parrot feather Myriophyllum aquaticum. Pest Management Science: formerly Pesticide Science 2007, 63, 798-802.

Dettenmaier E. Measuring and Modeling of Plant Root Uptake of Organic Chemicals. 2008, All Graduate Theses and Dissertations. 18. https://digitalcommons.usu.edu/etd/18.

Dettenmaier E, Doucette WJ. Mineralization and Plant Uptake of 14c-Labeled Nonylphenol, Nonylphenol Tetraethoxylate, and Nonylphenol Nonylethoxylate in Biosolids/Soil Systems Planted with Crested Wheatgrass. Environmental Toxicology and Chemistry 2007, 26, 193-200.

Dettenmaier E, Doucette W, Bugbee B. Chemical hydrophobicity and uptake by plant roots. Environmental Science & Technology 2009, 43, 324-329.

Doucette WJ, Shunthirasingham C, Dettenmaier EM, Zaleski RT, Fantke P, Arnot JA. A Review of Measured Bioaccumulation Data on Terrestrial Plants for Organic Chemicals: Metrics, Variability, and the Need for Standardized Measurement Protocols. Environmental Toxicology and Chemistry 2018, 37, 21-33.

Eggen T, Heimstad ES, Stuanes AO, Norli HR. Uptake and translocation of organophosphates and other emerging contaminants in food and forage crops. Environmental Science and Pollution Research 2013, 20, 4520-4531.

EPA, Ecological Effects Test Guidelines OCSPP 850.4800: Plant Uptake and Translocation Test, 2012.

Fantke P, Arnot JA, Doucette WJ. Improving plant bioaccumulation science through consistent reporting of experimental data. Journal of environmental management 2016, 181, 374-384.

Fredrickson DR, Shea PJ. Effect of soil pH on degradation, movement, and plant uptake of chlorsulfuron. Weed Science 1986, 34, 328-332.

Gobas FA, Burkhard LP, Doucette WJ, Sappington KG, Verbruggen EM, Hope BK, et al. Review of existing terrestrial bioaccumulation models and terrestrial bioaccumulation modeling needs for organic chemicals. Integrated environmental assessment and management 2016, 12, 123-134.

Grundmann S, Doerfler U, Munch JC, Ruth B, Schroll R. Impact of soil water regime on degradation and plant uptake behaviour of the herbicide isoproturon in different soil types. Chemosphere 2011, 82, 1461-1467.

Guha S, Mishra N. Clustering data streams. Data Stream Management. Springer, 2016, pp. 169-187.

Haykin SS. Neural networks and learning machines. Vol 3: Pearson Upper Saddle River, NJ, USA, 2009.

Hsu FC, Marxmiller RL, Yang AY. Study of root uptake and xylem translocation of cinmethylin and related compounds in detopped soybean roots using a pressure chamber technique. Plant Physiology 1990, 93, 1573-1578.

Kärrman A, Ericson I, van Bavel B, Darnerud PO, Aune M, Glynn A, et al. Exposure of perfluorinated chemicals through lactation: levels of matched human milk and serum and a temporal trend, 1996–2004, in Sweden. Environmental Health Perspectives 2007, 115, 226-230.

Kaur A, Kaur A. Comparison of fuzzy logic and neuro-fuzzy algorithms for air conditioning system. International journal of soft computing and engineering 2012, 2, 417-20.

Korf RE. Artificial intelligence search algorithms: Computer Science Department, University of California, 1996.

Limmer MA, Burken, J. G., Plant Translocation of Organic Compounds: Molecular and Physicochemical Predictors. Environmental Science & Technology Letters 2014, 1, 156-161.

Lipinski CA, Lead- and Drug-Like Compounds: The Rule-of-Five Revolution. Drug Discovery Today: Technologies 2004, 1, 337-341.

Martínez F, Lazo Y, Fernández-Galiano R, Merino J. Chemical composition and construction cost for roots of Mediterranean trees, shrub species and grassland communities. Plant, Cell & Environment 2002, 25, 601-608.

McBride MB, Shayler HA, Spliethoff HM, Mitchell RG, Marquez-Bravo LG, Ferenz GS, et al. Concentrations of lead, cadmium and barium in urban garden-grown vegetables: the impact of soil variables. Environmental Pollution 2014, 194, 254-261.

McGuinness M, Dowling D. Plant-Associated Bacterial Degradation of Toxic Organic Compounds in Soil. International Journal of Environmental Research and Public Health 2009, 6, 2226-2247.

McKone TE, Maddalena RL. Plant uptake of organic pollutants from soil: bioconcentration estimates based on models and experiments. Environmental Toxicology and Chemistry: An International Journal 2007, 26, 2494-2504.

Mitchell R, Michalski J, Carbonell T. An artificial intelligence approach: Springer, 2013.

Picó Y, Alvarez-Ruiz R, Alfarhan AH, El-Sheikh MA, Alobaid SM, Barceló D. Uptake and accumulation of emerging contaminants in soil and plant treated with wastewater under real-world environmental conditions in the Al Hayer area (Saudi Arabia). Science of the Total Environment 2019, 652, 562-572.

Nair DN, Padmavathy S. Impact of Endophytic Microorganisms on Plants, Environment and Humans. The Scientific World Journal 2014, 250693. https://doi.org/10.1155/2014/250693.

Pullagurala VLR, Rawat S, Adisa IO, Hernandez-Viezcas JA, Peralta-Videa JR, Gardea-Torresdey JL. Plant uptake and translocation of contaminants of emerging concern in soil. Science of the Total Environment 2018, 636, 1585-1596.

Pussemier L. Model calculations and measurements of uptake and translocation of carbamates by bean plants. Chemosphere 1991, 22, 327-339.

Rein A, Legind CN, Trapp S. New concepts for dynamic plant uptake models. SAR and QSAR in Environmental Research 2011, 22, 191-215.

Robinson BH, Mills TM, Petit D, Fung LE, Green SR, Clothier BE. Natural and induced cadmium-accumulation in poplar and willow: Implications for phytoremediation. Plant and Soil 2000, 227, 301-306.

Russell SJ, Norvig P. Artificial intelligence: a modern approach: Malaysia; Pearson Education Limited, 2016.

Schriever C, Lamshoeft M. Lipophilicity matters – A new look at experimental plant uptake data from literature. Science of The Total Environment 2020, 713, 136667.

Sharifan H, Bagheri M, Wang D, Burken JG, Higgins CP, Liang L, Liu J, Schaefer CE, Blotevogel J. Fate and transport of per-and polyfluoroalkyl substances (PFASs) in the vadose zone. Science of The Total Environment 2021, 771, 145427.

Shahid M, Dumat C, Khalid S, Schreck E, Xiong T, Niazi NK. Foliar heavy metal uptake, toxicity and detoxification in plants: A comparison of foliar and root metal uptake. Journal of hazardous materials 2017, 325, 36-58.

Shone M, Bartlett B, Wood AV. A comparison of the uptake and translocation of some organic herbicides and a systemic fungicide by barley: II. Relationship between uptake by roots and translocation to shoots. Journal of Experimental Botany 1974, 25, 401-409.

Sicbaldi F, Sacchi GA, Trevisan M, Del Re AA. Root uptake and xylem translocation of pesticides from different chemical classes. Pest Management Science 1997, 50, 111-119.

Tan WW, Chua TW. Uncertain rule-based fuzzy logic systems: introduction and new directions (Mendel, JM; 2001)[book review]. IEEE Computational Intelligence Magazine 2007, 2, 72-73.

Teng Y, Wu J, Lu S, Wang Y, Jiao X, Song L. Soil and soil environmental quality monitoring in China: a review. Environment international 2014, 69, 177-199.

Trapp S. Plant uptake and transport models for neutral and ionic chemicals. Environmental Science and Pollution Research 2004, 11, 33-39.

Trapp S. Fruit tree model for uptake of organic compounds from soil and air. SAR and QSAR in Environmental Research 2007, 18, 367-387.

Wen B, Wu Y, Zhang H, Liu Y, Hu X, Huang H, et al. The roles of protein and lipid in the accumulation and distribution of perfluorooctane sulfonate (PFOS) and perfluorooctanoate (PFOA) in plants grown in biosolids-amended soils. Environmental pollution 2016, 216, 682-688.

Wilson H, Boots B, Millward AA. A comparison of hierarchical and partitional clustering techniques for multispectral image classification. IEEE International Geoscience and Remote Sensing Symposium. 3. IEEE, 2002, pp. 1624-1626.

Zadeh LA. Fuzzy logic. Computer 1988, 21, 83-93.

Zhang M, Zhu L. Sorption of Polycyclic Aromatic Hydrocarbons to Carbohydrates and Lipids of Ryegrass Root and Implications for a Sorption Prediction Model. Environmental Science & Technology 2009, 43, 2740-2745.

VITA

The professional career of Majid Bagheri in Civil Engineering started in 2005 with the beginning of Associate's degree program at the Technical and Vocational College of Tehran–Enghelab Eslami in Iran. After two years, he enrolled in Bachelor's degree program and received my Bachelor in Civil Engineering (Construction) at Technical and Vocational College of Kermanshah in Iran. Due to the interest in Environmental Engineering, he received his Master's degree at K. N. Toosi University of Technology, Tehran, Iran in 2013. During this period, he worked on different problems in environmental engineering. With such experience, he started his PhD in Civil Engineering (Environmental emphasis) at Missouri University of Science and Technology in May 2017. After working on various research problems for four years, he received his PhD in Civil Engineering (Environmental emphasis) at Missouri University of Science and Technology in May 2021.